Transactions of the ASME

Transactions

of The American Society of Mechanical Engineers

Published on the tenth of every month, except March, June, September, and December

Boundary Layer Along Annular Walls in a Swirling Flow

By HSUAN YEH,[1] PHILADELPHIA, PA.

The development of incompressible turbulent boundary layers along concave and convex stationary annular walls is investigated analytically and experimentally for a swirling flow; i.e., flow with both tangential and axial mean velocities. It was found that the integral momentum equation describing this type of boundary layer contains three correction terms to the conventional equation. The combined influence of these corrections appears to promote the growth of the boundary layer next to a concave wall. Other differences between the boundary layers with swirl and those without swirl are pointed out and interpreted. Measurements on turbulence intensities appear to confirm such interpretations.

NOMENCLATURE

The following nomenclature is used in the paper:

C_f = local friction coefficient = $\tau_w \Big/ \dfrac{1}{2} \rho V_1^2$

$C_{fz} = \tau_{wz} \Big/ \dfrac{1}{2} \rho V_{z1}^2$

p = mean (time-averaged) pressure

r = radius, radial component of cylindrical co-ordinates

v_r, v_ϕ, v_z = fluctuating velocity along r, ϕ, z-directions, respectively

V_r, V_ϕ, V_z = mean (time-averaged) velocity along r, ϕ, z-directions, respectively

$V_* = (\tau_w/\rho)^{1/2}$

z = axial component of cylindrical co-ordinates

δ = boundary-layer thickness, determined by intercept of a power law with free-stream velocity when plotted on log-log scale.

δ^* = displacement thickness, defined by Equation [7]

θ_z, θ_ϕ = momentum thickness, defined by Equations [6] and [15]

τ_z, τ_ϕ = shear stresses along axial or tangential directions

$\tau_{wz}, \tau_{\phi w}$ = wall shear stresses along axial or tangential directions

ϕ = tangential component of cylindrical co-ordinates

Subscript 1 refers to the outer edge of boundary layer. A bar on top means time-average.

1 INTRODUCTION

This paper presents an investigation on the development of incompressible, turbulent boundary layer along annular walls when the main flow is both axial and tangential. A preliminary report on this investigation was presented in 1955 (1).[2]

Although extensive research has been made in recent years on turbulent shear flows, the majority of such work is limited to unidirectional flows; i.e., flows in which the mean velocity does not change appreciably in direction. Many interesting cases in practice fall into this class, such as channel flow, pipe flow, wake, boundary layer, and jet. However, there are important cases in which the mean flow is continuously changing in direction. This is the flow, for example, occurring in turbomachinery.

Recent detail investigations (2, 3) on fully developed turbulent flows in a curved channel, a problem first investigated by Wattendorf in 1935 (4), show that not only the mean velocity but also the turbulence characteristics are changed significantly by the curvature of the mean flow. An attempt has been made in (2) to explain these results on the basis of turbulent-energy production. (For a brief summary see Part 3 of the present paper.) Another recent investigation (5) on the influence of solid-body rotation on screen-produced turbulence tends to confirm the theory. Now the present investigation can be interpreted in view of these previous findings. Together with them, it serves to throw some light on the influence on turbulent flow due to a swirling mean velocity.

2 EXPERIMENTAL FACILITIES

The experimental tunnel consists of the following (Fig. 1):

(a) A vertical test section, 60 in. long, in the form of an annular space bounded by two concentric, cylindrical walls. The radii of the inner and outer walls are 5 in. and 10 in., respectively. The direction of flow in the test section is spirally upward.

(b) An inlet section which turns the initially radial inflow into the axial direction. The air enters radially to meet the inlet guide vanes. These are spaced like a ring at a diameter of 50 in. By turning these guide vanes to a proper angle, a prescribed amount of tangential velocity can be produced. The air then passes through a 90-deg bend so that the radial flow is turned to axial, but the angular momentum due to the tangential velocity is essentially unaffected. A unique feature of this inlet device is that the main body of the flow has constant angular momentum and constant total energy. The tangential-velocity distribution in the test section therefore follows that of a free vortex, except of course in the regions of boundary layer next to the walls.

(c) An axial-flow impeller and its driving motor, located downstream of and vertically above the test section. Running at 1500 rpm, the impeller draws the air and thereby maintains the flow.

Fig. 2 shows a detail view of the 60-in. test section. Seven measuring stations, numbered consecutively 1 to 7 in the direction of flow, are provided. The distance between each measuring station is 10 in. However, in order to limit the amount of experimental work, detailed measurements were made only at stations 3, 5, and 7. In this paper only the results at stations 3 and 7 are presented, these being considered sufficient to show the nature of the problem.

The instrumentation used in this investigation consisted of conventional velocity and direction probes for mean flow measurements and a hot-wire anemometer for turbulence measurements. The hot-wire anemometer is of the constant-current type, with full compensation up to 10,000 cps.
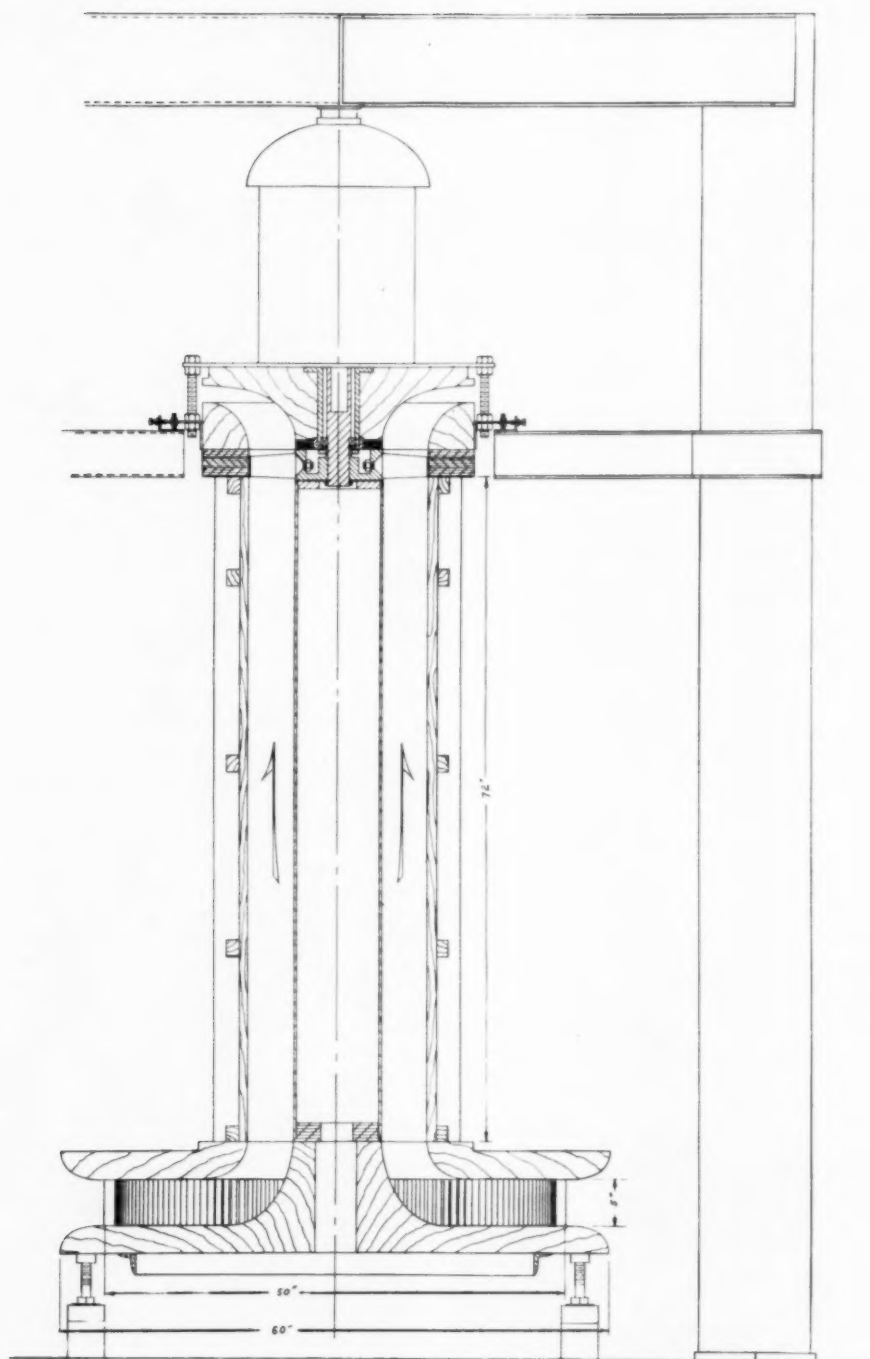
Fig. 1 Testing Setup Assembly
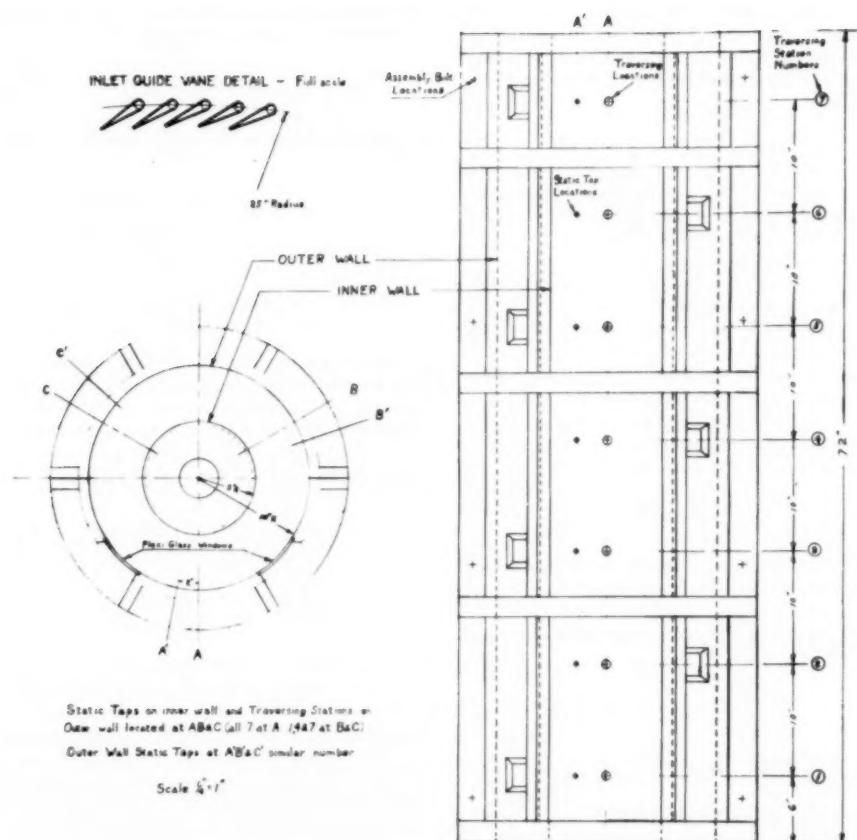
FIG. 2   DETAIL OF TEST SECTION

3   ANALYTICAL CONSIDERATIONS

Growth of Boundary Layer for Mixed Axial and Tangential Flows. In the following analysis we shall make the assumption of axial symmetry; i.e., $\partial/\partial\phi = 0$. The equations of motion in the cylindrical co-ordinate system then become:

Along the radial direction

$$V_r \frac{\partial V_r}{\partial r} + V_z \frac{\partial V_r}{\partial z} - \frac{V_\phi^2}{r} = -\frac{1}{\rho}\frac{\partial p}{\partial r} - \frac{\partial}{\partial r}(\overline{v_r^2})$$

$$-\frac{\partial}{\partial z}(\overline{v_r v_z}) - \frac{\overline{v_r^2}}{r} + \frac{\overline{v_\phi^2}}{r} + \nu\left(\nabla^2 V_r - \frac{V_r}{r^2}\right)\dots\dots[1]$$

Along the tangential direction

$$V_r \frac{\partial V_\phi}{\partial r} + V_z \frac{\partial V_\phi}{\partial z} + \frac{V_r V_\phi}{r} = -\frac{\partial}{\partial r}(\overline{v_r v_\phi}) - \frac{\partial}{\partial z}(\overline{v_\phi v_z})$$

$$-\frac{2\overline{v_r v_\phi}}{r} + \nu\left(\nabla^2 V_\phi - \frac{V_\phi}{r^2}\right)\dots\dots[2]$$

Along the axial direction

$$V_r \frac{\partial V_z}{\partial r} + V_z \frac{\partial V_z}{\partial z} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - \frac{\partial}{\partial r}(\overline{v_r v_z})$$

$$-\frac{\partial}{\partial z}(\overline{v_z^2}) - \frac{\overline{v_r v_z}}{r} + \nu\nabla^2 V_z\dots\dots[3]$$

where

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{\partial^2}{\partial z^2}$$

The equation of continuity is

$$\frac{1}{r}\frac{\partial}{\partial r}(rV_r) + \frac{\partial V_z}{\partial z} = 0$$

In the boundary layer $\frac{\partial}{\partial z} \ll \frac{\partial}{\partial r}$; hence we may write

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r}\dots\dots\dots[4]$$

It should be noted that Equation [4] does not necessarily mean that the boundary-layer thickness should be small compared to the radius of the wall. Furthermore, the following analysis does not make the usual assumptions of small turbulence and constant pressure in the direction normal to the boundary layer. In a turbomachine, the boundary layer builds up with the axial distance of travel and often occupies a considerable portion of the annular space of the flow. Because of the tangential mean velocity, neither the turbulence level nor the radial-pressure variation need be small in the boundary layer. The following derivation is still valid for such cases.

We shall now form the integral of the equations of motion in the following manner:

(a) Multiplying Equation [3] by r and integrating across the boundary layer from $r = r_w$ at the wall to $r = r_1$ at the outer edge of the boundary layer, we obtain

$$\int_{r_w}^{r_1} \left[\frac{\partial}{\partial r} (rV_rV_z) + \frac{\partial}{\partial z} (rV_z^2) \right] dr = -\int_{r_w}^{r_1} \frac{1}{\rho} \frac{\partial p}{\partial z} r dr$$
$$+ \frac{1}{\rho} [r\tau_z]_{r_w}^{r_1} - \int_{r_w}^{r_1} \frac{\partial}{\partial z} (\overline{v_z^2}) r dr \dots \dots [5]$$

where

$$\tau_z = -\rho \overline{v_r v_z} + \mu \frac{\partial V_z}{\partial r}$$

is the shear stress in the axial direction. The left-hand side of Equation [5], however, is equal to

$$r_1 V_{r1} V_{z1} + \int_{r_w}^{r_1} \frac{\partial}{\partial z} (rV_z^2) dr$$
$$= \frac{d}{dz} \int_{r_w}^{r_1} (V_z^2 - V_z V_{z1}) r dr + \frac{dV_{z1}}{dz} \int_{r_w}^{r_1} V_z r dr$$

We now define a momentum thickness θ_z and a displacement thickness δ^* for the flow in the axial direction such that

$$\theta_z = \frac{1}{r_w V_{z1}^2} \int_{r_w}^{r_1} V_z(V_{z1} - V_z) r dr \dots \dots [6]$$

$$\delta^* = \frac{1}{r_w V_{z1}} \int_{r_w}^{r_1} (V_{z1} - V_z) r dr \dots \dots [7]$$

Then the left side of Equation [5] becomes

$$-r_w V_{z1}^2 \frac{d\theta_z}{dz} - r_w V_{z1} \frac{dV_{z1}}{dz} (\delta^* + 2\theta_z) + V_{z1} \frac{dV_{z1}}{dz} \int_{r_w}^{r_1} r dr$$

Hence Equation [5] becomes, after setting $\tau_z = 0$ at $r = r_1$

$$\frac{\tau_{zw}}{\rho V_{z1}^2} = \frac{d\theta_z}{dz} + \frac{1}{V_{z1}} \frac{dV_{z1}}{dz} (\delta^* + 2\theta_z)$$
$$- \frac{1}{r_w V_{z1}^2} \frac{d}{dz} \int_{r_w}^{r_1} \left(\frac{p_1}{\rho} + \frac{V_{z1}^2}{2} \right) r dr$$
$$+ \frac{1}{\rho r_w V_{z1}^2} \frac{d}{dz} \int_{r_w}^{r_1} (p_1 - p) r dr - \frac{1}{r_w V_{z1}^2} \int_{r_w}^{r_1} \frac{\partial}{\partial z} (\overline{v_z^2}) r dr \dots [8]$$

This equation describes the growth along the axial direction of the boundary layer of the axial flow. It can be seen that with the exception of the last three terms, Equation [8] is identical to the well-known Karman integral momentum equation for the boundary layer along a flat plate. In the case of unidirectional flows, $(p_1/\rho + V_{z1}^2/2)$ is a constant and hence the third term on the right side of Equation [8] drops out. If the pressure is assumed constant across the boundary layer such that $p = p_1$, the fourth term on the right vanishes. Furthermore, the last term is usually also neglected. In the present case of swirling flow, however, none of these three terms can be assumed small without experimental evidence.

To evaluate the integral involving $(p_1 - p)$, we make use of Equation [1]. By a consideration of the order of magnitudes; viz., $V_r \ll V_\phi \approx V_z$ and $\partial/\partial z \ll \partial/\partial r$ we may write

$$\frac{1}{\rho} \frac{\partial p}{\partial r} = \frac{V_\phi^2}{r} - \frac{\overline{v_r^2}}{r} - \frac{\partial}{\partial r} (\overline{v_r^2}) + \frac{\overline{v_\phi^2}}{r} \dots \dots [9]$$

(In line with all previous derivations, turbulence quantities are not assumed to be small *a priori*.) Hence

$$\frac{p_1 - p}{\rho} = \int_r^{r_1} \frac{V_\phi^2}{r} dr + \int_r^{r_1} \frac{\overline{v_\phi^2} - \overline{v_r^2}}{r} dr + \overline{v_r^2} \dots [10]$$

Let the Bernoulli's constant of the flow outside the boundary layer be H_1

$$H_1 = \frac{p_1}{\rho} + \frac{V_{z1}^2 + V_{\phi1}^2}{2} \dots \dots \dots [11]$$

then

$$\frac{d}{dz} \left(\frac{p_1}{\rho} + \frac{V_{z1}^2}{2} \right) = \frac{d}{dz} \left(H_1 - \frac{V_{\phi1}^2}{2} \right) \dots \dots [12]$$

In the present investigation H_1 is a constant not only along the free streamlines but also across them, hence $dH_1/dz = 0$. This, however, need not always be the case in turbomachinery. Substituting Equations [10] and [12] into Equation [8], the result is

$$\frac{\tau_{zw}}{\rho V_{z1}^2} = \frac{d\theta_z}{dz} + \frac{1}{V_{z1}} \frac{dV_{z1}}{dz} (\delta^* + 2\theta_z)$$
$$+ \frac{1}{r_w V_{z1}^2} \int_{r_w}^{r_1} \frac{\partial}{\partial z} \left[\int_r^{r_1} \frac{V_\phi^2}{r} dr + \frac{V_{\phi1}^2}{2} \right] r dr$$
$$+ \frac{1}{r_w V_{z1}^2} \int_{r_w}^{r_1} \frac{\partial}{\partial z} (\overline{v_r^2} - \overline{v_z^2}) r dr$$
$$+ \frac{1}{r_w V_{z1}^2} \int_{r_w}^{r_1} \frac{\partial}{\partial z} \left[\int_r^{r_1} \frac{\overline{v_\phi^2} - \overline{v_r^2}}{r} dr \right] r dr \dots \dots [13]$$

Again, the last three terms on the right side constitute the modification to Karman's integral momentum equation for a flat plate. Among these three terms, the first arises due to the tangential mean velocity. The remaining two terms exist not just due to turbulence, but rather to the anisotropy of the turbulence intensities.[3]

Multiplying Equation [2] by r^2 and integrating across the boundary layer from $r = r_w$ to $r = r_1$, we obtain

$$\int_{r_w}^{r_1} \left[\frac{\partial}{\partial r} (r^2 V_r V_\phi) + \frac{\partial}{\partial z} (r^2 V_z V_\phi) \right] dr$$
$$= \frac{1}{\rho} [r^2 \tau_\phi]_{r_w}^{r_1} - \frac{\partial}{\partial z} (r^2 \overline{v_\phi v_z}) \dots \dots [14]$$

where
$$\tau_\phi = -\rho \overline{v_r v_\phi} + \mu r \frac{\partial}{\partial r} \left(\frac{V_\phi}{r} \right)$$

is the shear stress in the tangential direction. The left side of Equation [14] is equal to

$$r_1^2 V_{r1} V_{\phi1} + \int_{r_w}^{r_1} \frac{\partial}{\partial z} (r^2 V_\phi V_z) dr$$
$$= \frac{d}{dz} \int_{r_w}^{r_1} V_z(rV_\phi - r_1 V_{\phi1}) r dr + \frac{d(r_1 V_{\phi1})}{dz} \int_{r_w}^{r_1} V_z r dr$$

We now define a momentum thickness θ_ϕ for the flow in the tangential direction such that

$$\theta_\phi = \frac{1}{r_w V_{z1}(r_1 V_{\phi1})} \int_{r_w}^{r_1} V_z(r_1 V_{\phi1} - rV_\phi) r dr \dots \dots [15]$$

---

[3] A similar term, corresponding to the fourth term on the right side of Equation [13], appears also in unidirectional flows and was pointed out by Ross (6).

It is noted that $(r_1 V_{\phi 1} - r V_{\phi})$ is the angular-momentum defect. Then the left side of Equation [14] becomes

$$-r_w V_{z1}(r_1 V_{\phi 1}) \frac{d\theta_\phi}{dz} - r_w V_{z1} \frac{d(r_1 V_{\phi 1})}{dz} (\theta_\phi + \delta^*)$$

$$- r_w (r_1 V_{\phi 1})\theta_\phi \frac{dV_{z1}}{dz} + V_{z1} \frac{d(r_1 V_{\phi 1})}{dz} \int_{r_w}^{r_1} r dr$$

Hence Equation [14] becomes

$$\frac{r_w \tau_{\phi w}}{\rho V_{z1}(r_1 V_{\phi 1})} = \frac{d\theta_\phi}{dz} + \frac{1}{r_1 V_{\phi 1}} \frac{d(r_1 V_{\phi 1})}{dz} (\theta_\phi + \delta^*)$$

$$+ \frac{1}{V_{z1}} \left(\frac{dV_{z1}}{dz} \right) \theta_\phi + \frac{r_1{}^2 - r_w{}^2}{2 r_w(r_1 V_{\phi 1})} \frac{d(r_1 V_{\phi 1})}{dz}$$

$$- \frac{1}{r_w V_{z1}(r_1 V_{\phi 1})} \int_{r_w}^{r_1} \frac{\partial}{\partial z} (\overline{v_\phi v_z}) r^2 dr \dots \dots [16]$$

In the present investigation $r_1 V_{\phi 1}$ is a constant since the flow outside of the boundary layer possesses free-vortex type of tangential velocity. Hence Equation [16] is reduced for the present case to

$$\frac{r_w \tau_{\phi w}}{\rho V_{z1}(r_1 V_{\phi 1})} = \frac{d\theta_z}{dz} + \frac{1}{V_{z1}} \left(\frac{dV_{z1}}{dz} \right) \theta_\phi$$

$$- \frac{1}{r_w V_{z1}(r_1 V_{\phi 1})} \int_{r_w}^{r_1} \frac{\partial}{\partial z} (\overline{v_\phi v_z}) r^2 dr \dots \dots [17]$$

Influence of Swirling Flow on Turbulence. In interpreting subsequent experimental results, it is relevant to review briefly the conclusions of references (2, 3) with respect to the influence of mean velocity curvature on turbulence. If one takes the instantaneous (i.e., nonsteady) Navier-Stokes' equations, multiplies them by the instantaneous velocities and averages with respect to time, one obtains a set of six equations[4] describing the six Reynolds' stresses $\overline{v_i v_j}$. Of these six equations the three with $i = j$ serve to give an energy balance of the three turbulence intensities $\overline{v_i{}^2}$. Such a balance is between the production of any one turbulence intensity, its diffusion by turbulent motion and by pressure gradients, its transport by mean velocity, its energy transfer with other turbulence intensities, and its dissipation by viscosity. When applied to a plane, *straight* channel flow, these equations show that only the longitudinal (i.e., streamwise) turbulence intensity is produced directly from the mean flow, whereas the transverse (i.e., perpendicular to wall) and spanwise (i.e., parallel to wall but normal to flow) turbulence intensities receive their energy in turn from the longitudinal component. Furthermore, the turbulent energy is produced equally near the two walls. When these equations are applied to a plane, *curved* channel flow with purely tangential mean velocities, the following conclusions on the influence of a swirling mean flow can be drawn: (a) Turbulent energy is produced directly from the mean flow not only on the longitudinal component $\overline{v_\phi{}^2}$ but also on the transverse component $\overline{v_r{}^2}$; (b) $\overline{v_r{}^2}$ is produced near the outer or concave wall (where the tangential velocity decreases with radius) but is *suppressed* near the inner or convex wall (where the tangential velocity increases with radius), resulting in a larger intensity for both $\overline{v_z{}^2}$ and $\overline{v_r{}^2}$ near the outer wall. Experiments on the intensity, the spectrum, and the scale of $\overline{v_r{}^2}$ show indeed that large-scale turbulence eddies "roam" radially back and forth in the outer half of the curved channel, but that such motion is very much reduced in the inner half. One consequence of this strong radial turbulent

[4] See reference (7) for the derivation of these equations in Cartesian co-ordinates and reference (8) in cylindrical co-ordinates.

motion near the outer wall is that the mean velocity immediately adjacent to the wall is "pulled taut," resulting in a larger velocity gradient and shear stress at the outer wall. These facts turn out to be important in the following discussions.

4 EXPERIMENTAL RESULTS AND DISCUSSIONS

The experimental results presented in this paper consist of measurements taken at measuring stations 3 and 7 (see Fig. 2 for locations of these stations). The experiments can be divided into two parts: Part A, when the inlet guide vanes are set radial and hence with no swirl of the mean flow; and Part B, when the inlet guide vanes are set at an angle of approximately 25 deg from the radial direction with consequent swirl of the mean flow.

Part A (No swirl). Fig. 3 is a plot of total velocity $V (= V_z$ for no swirl) against r at stations 3 and 7. The slightly thicker boundary layer at the outer wall (compared to that at the inner wall) most likely results from the larger velocities that the fluid near the outer wall possesses in the 90-deg bend from the inlet guide vanes to the test section, with consequent larger loss.

It has been well established that the boundary-layer profile in the vicinity of a wall follows a universal law

$$\frac{V}{V_*} = A \log_{10} \left(\frac{y V_*}{\nu} \right) + B \dots \dots \dots \dots [18]$$

for $y V_* /\nu$ between 30 and approximately 1000. The constants A and B are not yet definitely determined, but a reasonable figure is $A = 5.6$, $B = 4.9$. Using these values for A and B and making use of the fact that $V_*/V_1 = (C_f/2)^{1/2}$, Equation [18] can be rewritten

$$\frac{V}{V_1} = \left(\frac{C_f}{2} \right)^{1/2} \left[5.6 \log_{10} \left(\frac{y V_1}{\nu} \right) \right.$$

$$\left. + 5.6 \log_{10} \left(\frac{C_f}{2} \right)^{1/2} + 4.9 \right] \dots \dots [19]$$

Thus for each value of C_f, this equation gives a curve of V/V_1 versus $y V_1/\nu$, Fig. 4. As suggested by Clauser (9), this offers a means of estimating C_f by merely plotting the experimental values of V/V_1 versus $y V_1/\nu$ and picking the C_f that best fits the data. When this is done to the boundary layers with no mean swirl, the values of C_f and the corresponding values of V_* are as follows:

Inner wall: Station 3....$C_f = 0.0040$ $V_* = 3.62$ fps
            Station 7....$C_f = 0.0036$ $V_* = 3.54$ fps
Outer wall: Station 3....$C_f = 0.0036$ $V_* = 3.43$ fps
            Station 7....$C_f = 0.0032$ $V_* = 3.34$ fps
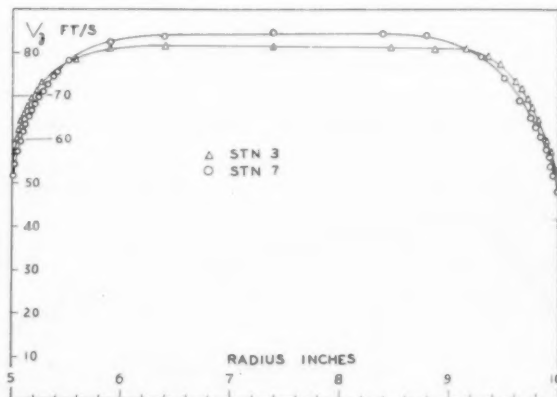
FIG. 3 AXIAL MEAN-VELOCITY DISTRIBUTION WITHOUT SWIRL
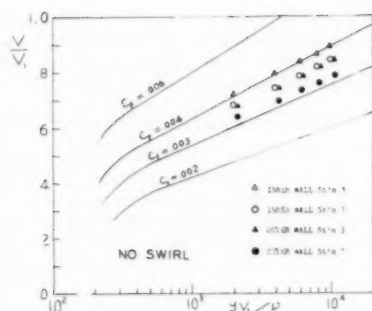
FIG. 4   MEAN-VELOCITY DISTRIBUTION IN VICINITY OF WALLS

FIG. 5   MEAN-VELOCITY DISTRIBUTION OF OUTER LAYER NO SWIRL

FIG. 6   AXIAL MEAN-VELOCITY DISTRIBUTION WITH SWIRL.

FIG. 7   TANGENTIAL MEAN-VELOCITY DISTRIBUTION WITH SWIRL
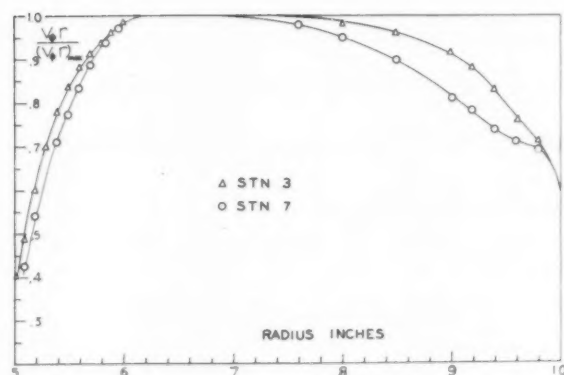
FIG. 8   ANGULAR MOMENTUM DISTRIBUTION WITH SWIRL
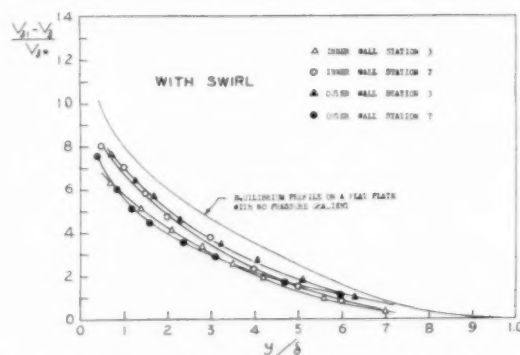
FIG. 9   MEAN AXIAL-VELOCITY DISTRIBUTION OF OUTER LAYER
WITH SWIRL
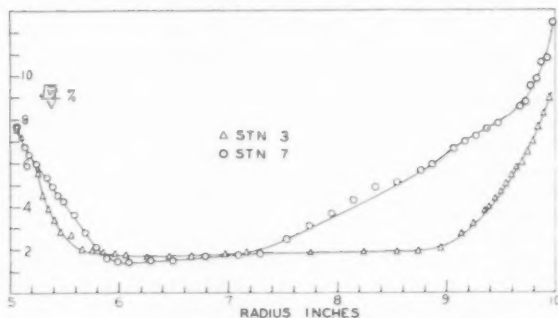
FIG. 10   DISTRIBUTION OF AXIAL-TURBULENCE INTENSITY WITH
SWIRL
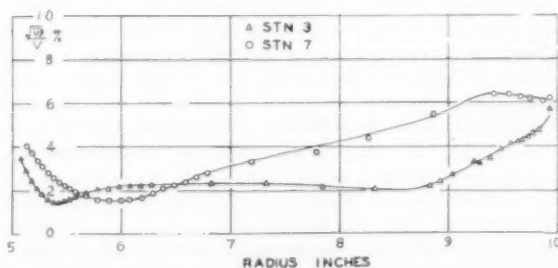
FIG. 11   DISTRIBUTION OF RADIAL-TURBULENCE INTENSITY WITH
SWIRL

These values of C_f are entirely reasonable. A value of $C_f = 0.0040$, for example, corresponds to a Reynolds number of 5×10^5 on a flat plate with zero pressure gradient in the direction of flow, equivalent to 1 ft from the leading edge for the present free-stream velocity. Furthermore, the decrease of C_f and V_* (the latter is a direct indication of the shear stress at wall) in the direction of flow behaves just like the boundary layer on a flat plate with zero pressure gradient.

The outer portion of the boundary layer can now be compared by plotting $(V_1 - V)/V_*$ against y/δ. As is well known, the shape of the velocity distribution in the outer layer, when plotted in this manner, is again a universal curve for all Reynolds numbers if the pressure gradient in the direction of flow is zero. In the present investigation, this pressure gradient is so small as to become practically negligible. The four boundary layers with no swirl, i.e., at inner and outer walls and at stations 3 and 7, are compared to the fully developed, equilibrium profile on a flat plate without pressure gradient in Fig. 5. It is clear that the boundary layers at station 3 are far from being in equilibrium, but that from station 3 to station 7 they gradually approach the equilibrium profile. Both the C_f-values and the profiles of Fig. 5 point out to the previously mentioned fact that the boundary layer at the outer wall is in a more developed stage than that at the inner wall, but that neither is fully developed since the flow is accelerating in the 90-deg bend just before entering the test section.

At this point it becomes clear that the drawbacks of the present experimental setup are (a) the different initial conditions at the two walls, and (b) the lack of an equilibrium boundary layer at the start of measurements. Unfortunately, these defects cannot be corrected easily, although future investigations should do well to take them into account. Since the present investigation is primarily intended to show the difference between a swirling flow and a flow without swirl, these defects are not particularly serious.

To summarize the character of growth of boundary layers without swirl: Both the C_f-values (an indication of the character of the close-to-wall layer) and the outer-layer profiles, Fig. 5, indicate a trend similar to the boundary-layer growth along a flat plate with no pressure gradient.

Part B (With Swirl). When the inlet guide vanes are turned to create in the test tunnel a tangential velocity of approximately 30 per cent (near outer wall) to 50 per cent (near inner wall) of the axial velocity, distinct changes are noticed in the character of the boundary layer next to the outer wall, although relatively little change is apparent next to the inner wall. The distribution of the axial mean 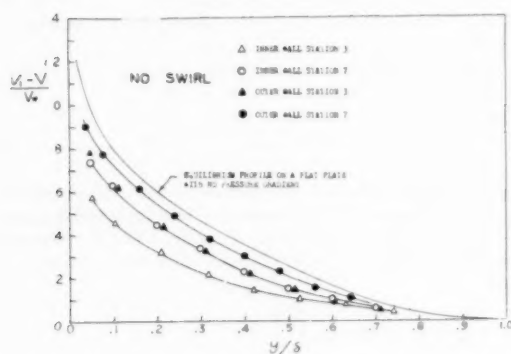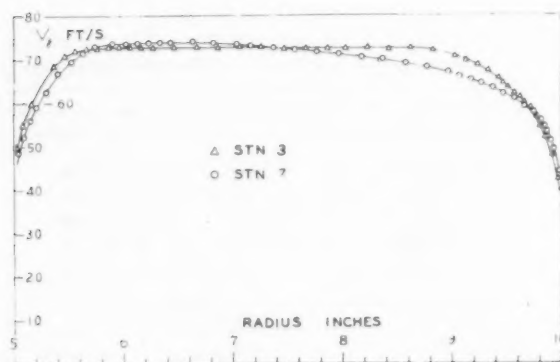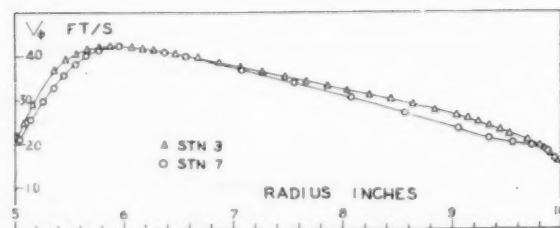velocity V_z and the tangential mean velocity V_ϕ at stations 3 and 7 are shown in Figs. 6 and 7. Instead of V_ϕ itself, it is perhaps more meaningful to consider the angular momentum $V_\phi r$, the distribution of which is shown in Fig. 8. From these figures and a comparison with Fig. 3 for no swirl, the following remarks can be made on the boundary layers at the outer wall with swirl: (a) They are thicker for both stations 3 and 7, but especially so for station 7, when compared to the case of no swirl; (b) in the immediate vicinity of the wall up to a distance of about 0.3 in., the mean velocity does not seem to decrease with distance, and in the case of V_z, actually seems to increase, in contrast to

usual boundary-layer growth; (c) farther out in the outer portion of the boundary layer, the edge of the boundary layer creeps rapidly into the free stream so that, at station 7, the boundary layer takes up a full 50 per cent of the annular space.

The foregoing qualitative description of the boundary-layer growth with swirl can be made more definitive by the following considerations: (a) From the V_z distribution we can evaluate C_{f_a} by a procedure similar to the case with no swirl. The results are as follows:

$$\text{Inner wall: Station 3} \ldots C_{f_a} = 0.0039$$
$$\text{Station 7} \ldots C_{f_a} = 0.0035$$
$$\text{Outer wall: Station 3} \ldots C_{f_a} = 0.0035$$
$$\text{Station 7} \ldots C_{f_a} = 0.0037$$

Thus in this respect the boundary layer at the inner wall behaves very much like that without swirl, but the boundary layer at the outer wall is quite different, its C_{f_a} does not seem to decrease with distance. Also, with these values of C_{f_a}, we can plot $(V_{z1} - V_z)/V_{z*}$ against y/δ, shown in Fig. 9. These curves indicate that, whereas the boundary-layer profile at the inner wall tends again to approach the equilibrium profile, that at the outer wall tends to depart more from it. (Similar conclusions can be drawn with respect to V_ϕ profile, but are omitted for brevity.)

The explanation of the peculiar behavior of the boundary-layer growth at the outer wall becomes quite clear when we consider the turbulence intensity distribution, Figs. 10 and 11. In Fig. 10 $(\overline{v_z{}^2})^{1/2}$, the rms of the axial turbulence velocity v_z, is plotted as a ratio of the total local mean velocity. In Fig. 11 $(\overline{v_r{}^2})^{1/2}$ the rms of the radial turbulence velocity v_r is plotted in the same way. These figures show that the turbulence intensity is generally larger near the outer wall. The radial intensity $(\overline{v_r{}^2})^{1/2}$ furthermore shows a dip near the inner wall, quite similar to the case of a curved channel of reference (2). As stated in the section Influence of Swirling Flow on Turbulence, the more intense turbulent motion near the outer wall tends to maintain a steep mean velocity gradient at the wall. At the same time, it brings about a more rapid contamination of boundary layer into the free stream.

One can now evaluate the various terms in Equation [13]. It develops that the last term of Equation [13] is negligibly small for all cases. The equation can now be written as follows

$$\frac{d\theta_s}{dz} = \frac{\tau_{zw}}{\rho V_{z1}{}^2} - \frac{1}{V_{z1}} \frac{dV_{z1}}{dz}(\delta^* + 2\theta_s)$$

$$\underbrace{- \frac{1}{r_w V_{z1}{}^2} \int_{r_w}^{r_1} \frac{\partial}{\partial z}\left[\int_r^{r_1} \frac{V_\phi{}^2}{r}\,dr + \frac{V_{\phi1}{}^2}{2}\right]r\,dr}_{\text{Correction I}}$$

$$\underbrace{+ \frac{1}{r_w V_{z1}{}^2} \int_{r_w}^{r_1} \frac{\partial}{\partial z}(\overline{v_z{}^2} - \overline{v_r{}^2})r\,dr}_{\text{Correction II}} \ldots \ldots [20]$$

Between stations 3 and 7, the various terms in Equation [20] are approximately as follows:

	$\dfrac{\tau_{zw}}{\rho V_{z1}{}^2}$	$\dfrac{-1}{V_{z1}} \dfrac{dV_{z1}}{dz}(\delta^* + 2\theta_z)$	Correction I	Correction II	$d\theta_z/dz$ from Eq. [20]
At inner wall...	0.0018	−0.0001	0	0	0.0017
At outer wall...	0.0018	−0.0001	−0.0001	+0.0002	0.0018

(The actual $d\theta_s/dz$ from measurements is about 20 per cent less at the inner wall and 20 per cent more at the outer wall, and it is suspected that the test tunnel may not have complete axial symmetry.) The combined influence of the correction terms in Equation [20] appears to promote the boundary-layer growth at the outer or concave wall for a swirling mean flow. These correction terms, however, do not tell the whole story. The influence of swirl is exerted even more strongly through the value of the wall shear stress. In this last respect, it again appears that the growth of boundary layer near a concave wall is promoted by a swirling mean flow.

It might be interesting to mention briefly a related experiment on the boundary-layer growth along *rotating* annular walls. In the experiment described in reference (5), both the flow and the annular walls were rotating. It was found that, in contrast to the case of swirling flow with stationary walls, the angular momentum decreases with radius near the inner wall and increases with radius near the outer wall. Based on the turbulent-energy production considerations, it can be shown that the radial turbulent motion is now promoted near the inner wall and suppressed near the outer wall. As reported in reference (5), the experiments indeed show that the boundary layer of the axial velocity near the inner wall grows more rapidly, and that near the outer wall less rapidly, when compared to the case of stationary walls with nonswirling flows.

Conclusions

For the present experimental setup and within the limited axial distance under survey, the following conclusions can be drawn for the incompressible boundary-layer growth of a swirling mean flow:

(a) The boundary-layer growth is governed by Equation [13] for the axial mean velocity and Equation [16] for the tangential mean velocity.

(b) Equation [13] differs from the conventional integral momentum equation in the appearance of three correction terms. Among these, one is found to be negligible. The remainder consists of one term resulting from tangential mean velocity and another term from the anisotropy of turbulence.

(c) The combined influence of the correction terms appears to promote the growth of the boundary layer at a stationary concave wall.

(d) The wall shear stress decreases with axial distance at a convex wall but appears to remain constant or even increase slightly at a concave wall, in both cases the pressure gradient in the direction of flow being negligibly small. Closely related to this phenomenon is the fact that the mean velocity in the immediate vicinity of a concave wall does not seem to slow down with distance of travel.

(e) In contrast to the behavior of the close-to-wall layer, the outer portion of the boundary layer next to a concave wall creeps rapidly into the free stream.

(f) The foregoing phenomena can be explained by the intense radial turbulent motion in the boundary layer next to a concave wall, as evidenced by the direct production of the radial turbulence intensity in the turbulent-energy equations for regions where the angular momentum decreases with radius.

Acknowledgments

BIBLIOGRAPHY

1 "Development of Boundary Layers of Mixed Tangential and Axial Flows in an Annular Space of Revolution," by H. Yeh, presented at the Diamond Jubilee Semi-Annual Meeting, Boston, Mass., June 19–23, 1955, of The American Society of Mechanical Engineers.

2 "An Investigation on Fully Developed Turbulent Flows in a Curved Channel," by S. Eskinazi and H. Yeh, *Journal of the Aeronautical Sciences*, vol. 23, no. 1, 1956, pp. 23–35.

3 "Further Investigations on Fully Developed Turbulent Flows in a Curved Channel," by H. Yeh, W. G. Rose, and H. Lien, final report to Office of Naval Research, Nonr-248(33), Mechanical Engineering Department, The Johns Hopkins University, Baltimore, Md., December, 1956.

4 "A Study of the Effect of Curvature on Fully Developed Turbulent Flow," by F. L. Wattendorf, Proceedings of the Royal Society of London, vol. 148, 1935, pp. 565–598.

5 "The Influence of Solid Body Rotation on Screen Produced Turbulence," by S. C. Traugott and H. Yeh, final report to National Advisory Committee on Aeronautics, Contract Naw 6408, September, 1956.

6 "Integration of the Reynolds Equations for Incompressible Turbulent Boundary Layers," by D. Ross, Ordnance Research Laboratory, Pennsylvania State College, May 29, 1953.

7 "The Structure of Turbulent Shear Flow," by A. A. Townsend, Cambridge University Press, Cambridge, England, 1956, p. 26.

8 "A Program of Research in the Field of Turbulent Flow in Ducts, in a Space of Revolution, and in Turbomachinery," by G. F. Wislicenus and H. Yeh, report to Office of Naval Research, Nonr 248(33), Mechanical Engineering Department, The Johns Hopkins University, Baltimore, Md., December, 1952.

9 "Turbulent Boundary Layers in Adverse Pressure Gradients," by F. H. Clauser, *Journal of the Aeronautical Sciences*, vol. 21, February, 1954, pp. 91–108.

Discussion

FRANK KREITH,[5] The experimental data and their analysis represent a valuable contribution to our understanding of the mechanism of turbulent flow. The swirling type of flow studied by the author has not only applications in the field of turbomachinery but occurs also in some heat-transfer equipment. In fact, the results reported by the author corroborate observations made at the Jet Propulsion Laboratory of the California Institute of Technology and at Lehigh University in connection with two different heat-transfer studies.[6,7]

In the first of these experiments, convection heat transfer from a convex surface was compared experimentally with heat transfer from a concave surface of the same radius of curvature. The geometry of the test section employed in this investigation corresponds closely to the curved channel used by the author in the work reviewed in Part 3 of his paper. It was found experimentally that at a given Reynolds number, the heat-transfer coefficient for an incompressible fluid flowing in a curved channel is substantially higher when the fluid is heated from a concave surface than when the fluid is heated from a convex surface. Since the rate of heat transfer by convection increases with increasing turbulence intensity and surface shear, the results of the

heat-transfer tests represent a graphic demonstration of the phenomena reported in this paper; namely, the transverse turbulent-velocity component v^2 is produced near a concave wall but is suppressed near a convex wall, a phenomenon which gives rise to a larger shear stress at the concave wall.

At the time the heat-transfer data were analyzed, the author's data were not available and Wattendorf's results (reference 4 of the Bibliography) were used instead. It was found that Wattendorf's data on the time-average velocity distribution could be correlated with reasonable accuracy by a power law in which the exponent, n, is a function of a dimensionless curvature parameter

$$\nu / [R_c(\tau_0/\rho)^{1/2}]$$

Since the range of the curvature radii used by Wattendorf was limited, it would be of interest if the author would compare his data from reference (3), which are not generally available and are only referred to qualitatively in the present paper, with those of Wattendorf and, if possible, relate empirically the shear stress and the velocity distribution near the wall to a generalized curvature parameter. This not only would make it possible to apply the author's result directly to a better interpretation and correlation of convective heat transfer in curved channels, but also could lead to an improvement in the evaluation of convective heat-transfer coefficients in convergent-divergent rocket nozzles. In one such study,[6] a 1/7-power law was used in the absence of better data irrespective of the wall curvature.

In another study,[7] the heat transfer to and from a fluid in swirling flow was investigated. In this study, a swirling motion was imparted to water and air flowing in a pipe by means of spiral strips or wires. It was found that the heat-transfer coefficients at the concave surface were as much as four times as large as corresponding coefficients in straight pipe flow. The increase in the heat-transfer coefficients was larger when the fluid was heated than when the fluid was cooled. It is believed that in the first case, the density gradient induces additional flow instabilities which promote turbulent mixing and heat transfer. Also these observations illustrate the phenomena described by the author and can be explained, at least qualitatively, in terms of the turbulence measurements reported in this paper.

The results of the heat-transfer experiments indicate that centrifugal forces affect the turbulent mixing process appreciably. Additional comments regarding the effect of centrifugal forces on the velocity profile and the shear stress would be a welcome addition to the author's discussion of the importance of the individual terms in his Equation [13].

L. H. SMITH, JR.[9] This paper is another fine contribution which the author has added to his previous works on turbulent flow in internal-flow systems (1-3, 5, 8 of his Bibliography). The problem is approached in all these studies through the Reynolds equations. This approach focuses attention on the fluctuating nature of the flow and permits a ready estimate of the effects which the fluctuations have on the equations of motion.

In turbomachinery the flow fluctuates in time not only because of "normal" turbulence but also because of the passing of rotor wakes. Root-mean-square fluctuations behind rotors as high as 15 per cent of the through-flow velocity are not uncommon near mid-annulus, and much greater magnitudes occur near the

annulus walls. Furthermore, the wakes dissipate rather rapidly,[10] especially near the blade trailing edges where the measurements must be taken in multistage turbomachinery, so that large derivatives with respect to z are present. This is significant since z-derivatives appear in the turbulence terms in Equations [13] and [17]. Perhaps the author would give us his opinion on the practicability of applying these equations to turbomachinery wall layers.

G. F. WISLICENUS.[11] This paper constitutes another valuable contribution to a series of investigations of curved turbulent flow[12] that have been conducted under the guidance of its author. It is hoped that these investigations will serve to stimulate the interest in phenomena which, although of great practical importance in internal fluid mechanics, previously may have been somewhat neglected in turbulence research.

The writer should like to make two observations of qualitative character in the hope that the author will render these thoughts usable by comparison with his analytical and experimental results.

The author's data seem to confirm in a general manner an old criterion by Prandtl regarding the stability of curved fluid motions. Prandtl reasoned that a curved flow with radially (outward) increasing angular momentum would be stable since a portion of the fluid departing radially from the mean flow would maintain its angular momentum and therefore would have a higher peripheral velocity than its surroundings when displaced radially inward, and a lower peripheral velocity than its surroundings when displaced radially outward. This would generate forces tending to return the displaced fluid portion to its original radius of travel. Inversely, a fluid motion with radially (outward) decreasing angular momentum would generate forces tending to move a radially displaced fluid portion farther away from its original radius of travel, thus rendering the flow unstable. It is evident that the boundary-layer flow near the inner, convex wall of the author's "vortex tunnel" will by this criterion be "stable" and near the outer, concave wall "unstable." At least the latter statement appears to be well supported by the author's repeated observation of increased turbulence near the outer, concave wall of a curved flow, and the generation of turbulence, indicated theoretically for this region, is thus in agreement with Prandtl's reasoning. In this case we are concerned with *strong* departures from a flow of uniform angular momentum. Already Wattendorf recognized the significance of this consideration regarding the question whether a turbulent flow of constant angular momentum or with "solid-body rotation" would have zero turbulent shear stress. This problem involves, of course, much smaller differences in velocity gradient than the boundary flow. The author's opinion about this problem would be of considerable interest.

The second observation of this writer has been suggested by Furgerson[13] who was not at the meeting at which the paper was presented. It concerns the behavior of the boundary layer of the axial-flow component, particularly its refusal to "grow" axially in the immediate neighborhood of the outer wall. Furgerson pointed out to the writer that the reduction of the peripheral component of the flow by friction would necessarily reduce the pressure difference between the inner and the outer walls.

[6] "An Approximate Solution of Compressible Turbulent Boundary-Layer Development and Convective Heat Transfer in Convergent-Divergent Nozzles," by D. R. Bartz, TRANS. ASME, vol. 77, 1955, pp. 1235-1245.

[9] Aircraft Gas Turbine Division, General Electric Company, Cincinnati, Ohio. Mem. ASME.

[10] "Low-Speed Wake Characteristics of Two-Dimensional Cascade and Isolated Airfoil Sections," by S. Lieblein and W. H. Roudebush, NACA TN 3771, October, 1956.

[11] Ordnance Research Laboratory, Pennsylvania State University, University Park, Pennsylvania. Mem. ASME.

[12] See Bibliography of the paper.

[13] W. T. Furgerson, Head Development Engineer, Oak Ridge National Laboratories, Oak Ridge, Tenn.

Thereby the pressure drop in the axial direction is increased along the outer (high-pressure) wall over and above the natural "friction pressure drop," thus tending to accelerate the axial component close to the outer wall as indicated by the author's results. For the same reason the axial pressure drop near the inner (convex) wall is reduced and may indeed lead to an axial pressure increase near that wall. Furgerson recognized the significance of this adverse pressure gradient with respect to separation.

The writer believes that the author's "Correction I" in his Equation [20] or [13] does indeed represent the effect which was qualitatively described by Furgerson, as it contains the axial gradient ($\partial/\partial z$) of an integral involving a radial pressure gradient (V_ϕ^2/r). It would be appreciated if the author would check this conclusion which, if true, may be useful by illustrating the physical meaning of the author's derivations. It should be observed that the author found this term to be of the same order as the second term in von Karman's integral momentum equation, with only a fairly slight swirling component. Therefore, this term might become dominating in more strongly swirling motions and may indeed lead to separation at the inner wall.

Author's Closure

The author wishes to express his sincere thanks to all the discussers for their valuable comments. It is indeed gratifying to learn of Professor Kreith's heat-transfer experiments which verify remarkably well the thesis of this paper, namely the different character of the turbulent boundary layers near a concave as against a convex wall. With respect to a power law to describe the time-average velocity distribution, the author feels that such a method is likely to be unsatisfactory to give a good representation in turbulent boundary layers. Even in the case of the simplest equilibrium profile, that for zero pressure gradient on a flat plate, the exponent of the power law varies with the Reynolds number from one seventh to one eighth or one ninth. With curved walls the spread of this exponent is likely to be much larger. For example, in the region near the outer wall in the test curved channel of reference (3) (which has an inner-to-outer wall-radii ratio of 0.75), the exponent is approximately $1/13$; whereas in the region near the inner wall of the same channel it is $1/5.6$. The first value does not seem to correlate the results of reference (4). It is the author's opinion that the velocity distribution can only be compared on the basis of Equation [18] for the region close to a wall, and of the form[14] $(v_1 - v)/v_* = f(y/\delta)$ for regions farther

[14] For tangential velocities it would be of the form $[(V_\phi r)_1 - V_\phi r]/r \left[\dfrac{1}{\rho} \dfrac{\partial p}{\partial \phi} \dfrac{be}{r_p}\right]^{1/2} = f(y/be)$ suggested by Wattendorf in reference (4).

out from a wall. Such a representation is less convenient in application than a power law, but appears to be more logical.

Dr. Smith brought out the interesting point that in turbomachinery the fluctuating velocities are not only due to turbulence but also due to the passing of rotor wakes. The latter type of fluctuating velocity would be periodic and would appear mainly as $\overline{v_\phi^2}$ and $\overline{v_z^2}$. Since their derivatives with respect to z may not be small, it is possible that the last term containing $(\overline{v_\phi^2} - \overline{v_r^2})$ in Equation [13], the term that is found small in this paper and neglected in Equation [20], may not be negligible in general in turbomachinery. Unfortunately there are no experimental data to the author's knowledge that can show the relative importance of this term, although the form of this term suggests a light promotion of boundary-layer growth. In contrast, the decay of the periodic $\overline{v_z^2}$ due to passing wakes would decrease the value of the Correction Term II in Equation [20], resulting in lesser tendency for boundary-layer growth. These statements are purely tentative in nature. It is quite possible that the less obvious influence of the periodic fluctuations on wall shear stress is the more dominant factor, overwhelming the more obvious "correction terms" of this paper.

Prandtl's stability criterion[15] of curved fluid motions described by Professor Wislicenus is indeed very pertinent to the problem of this paper. The author pointed out in reference (3) that this criterion is based on the sign of $d(V_\phi r)/dr$, whereas the direct production (or suppression) of $\overline{v_r^2}$ is based on the sign of $\overline{v_r v_\phi}$. Since $\overline{v_r v_\phi}$ usually changes sign when $d(V_\phi r)/dr$ does so, Prandtl's criterion essentially agrees with the consideration of turbulent energy equations. However, it must be noted in this respect that the place at which $\overline{v_r v_\phi}$ changes sign in a curved flow is not necessarily the place at which either $d(V_\phi r)/dr$ or $d(V_\phi/r)/dr$ changes sign. After all, the turbulence characteristic at a point is not only influenced by the mean velocity or its gradient at that point, but also by the flow conditions through the whole field as well.

Professor Wislicenus is entirely correct in identifying the Correction Term I of Equation [20] to the physical situation suggested by Mr. Furgerson. This term is always negative near a concave wall and positive near a convex wall. Although its value near the convex wall is found to be negligibly small in the present experiment, it can certainly be of significant magnitude if the swirling velocity is strong, leading to a rapid thickening of boundary layer and an early separation at the convex wall.

[15] It appears that Rayleigh first suggested this criterion in his paper, "On the Dynamics of Revolving Fluids," Proceedings of the Royal Society, vol. 6, series A, 1916, pp. 148–154.

Stall Propagation in a Cascade of Airfoils

By A. H. STENNING[1] and A. R. KRIEBEL[2]

A small perturbation analysis of rotating stall is presented which includes the effects of finite pressure rise across the cascade, finite blade chord, and boundary-layer time delay. The theory yields a stability criterion for the initiation of rotating stall, upper and lower boundaries for the velocity of propagation, and a possible explanation for the mechanism governing the number of stall cells Tests on rotating stall in a cascade and two rotors suggested a vortex-shedding model for fully developed rotating stall which is not limited to small perturbations and which appears to agree closely with the observed stalling mechanism.

NOMENCLATURE

The following nomenclature is used in the paper:

A = area
$a_n(t)$ = functions of time
$b_n(t)$ = functions of time
b = half-wave length of the disturbance
c = velocity relative to blade row
c_x = steady state velocity in x-direction
c_y = steady state velocity in y-direction
C_p = cascade pressure coefficient = $\dfrac{p_2 - p_1}{\rho/2c_1{}^2}$
D = operator $\partial/\partial t$
L = equivalent chord length of blade
n = number of harmonic
p = stream pressure
p_0 = total pressure
u = total velocity in x-direction
v = total velocity in y-direction
V = velocity of vortex downstream of cascade
V_p = velocity of stall propagation
V_i = induced velocity within stall cell
α = discharge coefficient of cascade
$\alpha' = \dfrac{d\alpha}{d(\cot \beta_1)}$
β_1 = inlet angle to cascade
β_2 = outlet angle from cascade
θ = angle of vortex street
δ = perturbation quantity
λ = root of characteristic equation
σ = cascade solidity $\dfrac{\text{chord}}{\text{pitch}}$
ω = radian frequency
τ = time constant of boundary-layer delay

[1] Assistant Professor of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Mass.
[2] Research Engineer, Stanford Research Institute, Menlo Park, Calif.

ρ = density
Φ = total velocity potential
φ = perturbation velocity potential
$\varphi_x = \dfrac{\partial \varphi}{\partial x}$
$\varphi_y = \dfrac{\partial \varphi}{\partial y}$
$\varphi_t = \dfrac{\partial \varphi}{\partial t}$

1 INTRODUCTION

When a diffusing cascade of airfoils is operated at a high angle of attack, a local disturbance may initiate stall on one of the airfoils. The stalled airfoil restricts the flow through the channel adjacent to its upper surface, and in consequence the fluid is deflected around the blocked channel as shown in Fig. 1, increasing the angle of attack on the blade above the stalled airfoil and decreasing the angle of attack on the blade below so that the disturbance propagates along the cascade. If the blades are arranged in a closed circle, or in an annulus (as in an axial compressor), one or more stall cells may appear, moving with steady velocity around the cascade. When the stall cell passes over a blade, the blade loading changes rapidly, and the resulting excitation may cause blade failure if the cascade is operated in rotating stall for any length of time

The vibration problems associated with rotating stall are of concern to the axial-compressor designer, who would much rather avoid rotating stall than design the blading to withstand it. If it were possible to predict the number and velocity of propagation of the stall cells in a compressor, the blades could possibly be designed to avoid dangerous resonance conditions. Failing this, if the number of stall cells could be changed by making small changes in the machine geometry, it would at least be possible to doctor a sick compressor. Unfortunately, neither of these objectives has yet been attained, and experimental investigations of rotating stall in compressors and cascades have yielded
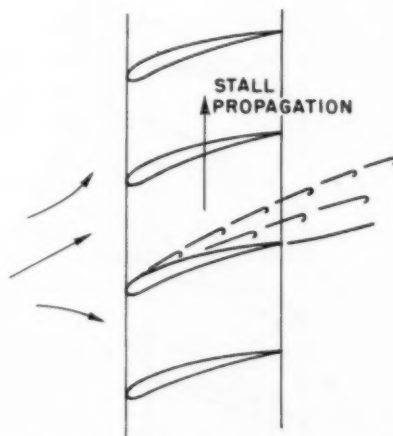

FIG. 1 EFFECT OF STALLED BLADE ON FLOW ENTERING CASCADE

a bewildering variety of results (1, 2, 3, 4)[3] which have not yet been unified by theory.

In order to study the phenomenon in its simplest form, a circular cascade tunnel was constructed at the Massachusetts Institute of Technology under the sponsorship of the National Advisory Committee for Aeronautics, with the objective of studying in detail the propagation of stall cells along a cascade. In addition, stall propagation was investigated in two rotating blade rows, with and without inlet guide vanes.

Early linearized analyses of the problem (1, 5, 6) considered the propagation of stall cells having wave lengths much greater than the blade chord and were therefore unable to explain the varying wave lengths of the stall cells found in compressors. In Section 2 a small perturbation analysis is presented which permits the inclusion of finite blade-chord effects and which yields the same results as the earlier analyses when applied to the cases for which they are valid. Photographic study of rotating stall in the circular cascade suggested a vortex-model representation which is not limited to small disturbances and which appears to describe the flow in the cascade much better than the linearized analyses. An analysis based upon this flow model is presented in Section 4.

2 SMALL PERTURBATION ANALYSIS

2.1 *Basic Assumptions.* In any analysis of stall propagation, the results obtained depend on the assumptions made about (a) the nature of the flow field after the cascade, (b) the form of the cascade characteristic, and (c) the dynamic response of the cascade to changes in inlet angle of attack.

The flow field immediately behind a stalled cascade in steady flow consists of streams of fluid which have suffered little loss in stagnation pressure, separated by regions of low stagnation pressure shown in Fig. 2. Downstream, mixing occurs between the high and low-velocity regions, and at a distance of ten chord lengths or so from the cascade the velocity is approximately uniform across the wake. In analyzing stall propagation, some approximation to the real conditions after the cascade in the unsteady flow must be made. Emphasis may be placed on the flow field immediately after the cascade by considering the flow to consist of a number of free jets entering a region of constant pressure. Alternatively, the region where mixing is complete

[3] Numbers in parentheses refer to the Bibliography at the end of the paper.
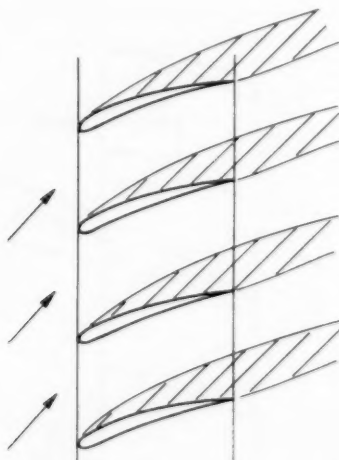
FIG. 2 FLOW FIELD OF STALLED CASCADE

may be considered to be of greater importance and the velocity distribution downstream assumed continuous throughout the field. The true condition lies somewhere between these two extremes, and in this investigation both of the extreme cases are considered so that boundaries of the possible solutions may be established.

The cascade performance may be represented in terms of pressure coefficient and outlet angle, or lift coefficient and drag coefficient, or effective outlet area and outlet angle, all expressed as functions of inlet angle to the cascade. In this analysis the cascade is represented by a series of channels in parallel, with variable-area outlets to reproduce the blockage effect of the separated flow. The effective outlet area is assumed to be a continuous function of the inlet angle in steady flow. The flow angle leaving the cascade is considered to be independent of inlet angle, a satisfactory assumption for closely spaced blades.

In unsteady flow, the cascade does not respond immediately to changes in inlet angle, and a finite time is required for the boundary layer to reach a new equilibrium condition after a change in angle of attack. The simplest way to represent a dynamic effect of this type is to approximate the response with a simple exponential time lag.

By examining the stability of small disturbances superimposed on the steady flow, the conditions which permit disturbances to propagate unchanged and the velocity of propagation will be determined.

2.2 *Representation of the Cascade.* The analysis is based on the model shown in Fig. 3, with the cascade simulated by a series
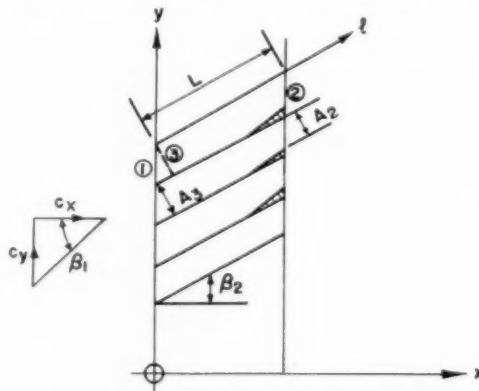
FIG. 3 CASCADE REPRESENTATION AND NOTATION

of channels of length L arranged in parallel, with variable-area outlets. The fluid enters the cascade at 1 with entrance angle β_1, is turned to β_2 at 3 in a short distance, and leaves the channel at 2.

The ratio of exit area to inlet area, A_2/A_3, is defined to be a function of β_1 when the flow is steady

$$\frac{A_2}{A_3} = \alpha = F(\cot \beta_1)$$

The assumption is made that local changes in α lag behind local changes in β_1 exponentially as shown in Fig. 4, so that for step changes in β_1

$$\delta\alpha = (\delta\alpha)_{ss}[1 - e^{-t/\tau}]$$

where $(\delta\alpha)_{ss}$ is the steady state change in α corresponding to the change in β_1 and τ is the time constant of the boundary-layer time delay. Thus, in general, when β_1 changes continually

$$\tau\frac{\partial(\delta\alpha)}{\partial t} = (\delta\alpha)_{ss} - \delta\alpha$$

at any instant where $(\delta\alpha)_{ss}$ is the steady-state value of $\delta\alpha$ corresponding to the value of β_1 at that instant or

$$\delta\alpha = \frac{(\delta\alpha)_{ss}}{\tau D + 1}$$

using operational notation where D denotes $\partial/\partial t$.

This representation of the boundary-layer response is an extreme oversimplification of a complex phenomenon which is not yet fully understood. It is, of course, not possible to consider the changes in effective exit area as independent of the fluid velocity in the passage since the fluid itself forms the "gate." However, the response assumed gives the simplest model which includes the boundary-layer delay and should at least yield some information on the effect of this delay.

A_2/A_3 is equivalent to the ratio

$$\frac{\text{actual flow through cascade}}{\text{ideal flow for same } (p_{01} - p_2) \text{ with no losses}}$$

and is obtainable, for a real cascade, from test results.

It can be seen that $\alpha = A_2/A_3$ is a measure of the "swallowing capacity" of the cascade and must therefore be important when the possibility of flow spillage around the entrance is considered. For a real cascade, it can be shown that α is equivalent to

$$\frac{\cos \beta_1}{\cos \beta_2 \sqrt{(1 - C_p)}}$$

for a rectilinear cascade, where C_p is the pressure coefficient (see Appendix).

Additional assumptions are that the fluid is incompressible and frictionless, and that changes in β_2 in the unsteady flow can be neglected. The blades are considered to be very close together, so that α may be taken as a continuous function of y.

For the initial analysis, the boundary-layer response is assumed fast compared with the inertia delays so that τ is taken as zero, and the case with constant outlet pressure is considered. The effects of finite τ and fluctuating back pressure are considered later.

2.3 *Solution for* $\tau = 0$ *With No Downstream Mixing.* The cascade lies on the y-axis in the x-y plane. The fluid is considered to be incompressible. Perturbations are considered from a steady flow with inlet angle β_1 and velocity components c_x, c_y.

Since the flow entering the cascade is irrotational a velocity potential can be used.

In the unsteady flow

$$u = c_x + \varphi_x$$
$$v = c_y + \varphi_y$$
$$\Phi = c_x x + c_y y + \varphi$$

where Φ is the total velocity potential and φ is the perturbation potential.
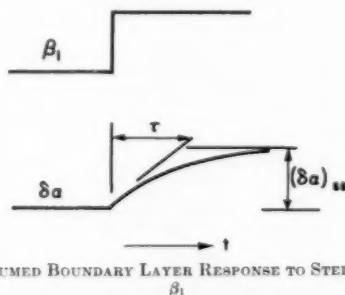
FIG. 4 ASSUMED BOUNDARY LAYER RESPONSE TO STEP CHANGES IN β_1

From continuity, $\partial u/\partial x + \partial v/\partial y = 0$; i.e. $\varphi_{xx} + \varphi_{yy} = 0$. Solutions to this equation are obtainable in the form

$$\varphi = \sum_{n=1}^{\infty} \left[a_n(t) \cos \frac{n\pi y}{b} + b_n(t) \sin \frac{n\pi y}{b} \right] e^{\frac{n\pi x}{b}} \quad \ldots [1]$$

The stability of the disturbances depends on the time-dependent functions $a_n(t)$, $b_n(t)$, which are as yet undetermined. From consideration of the dynamic equations, we may find the conditions required for the disturbances to be damped, amplified, or propagated unchanged along the cascade.

From Euler's equation for unsteady flow (7)

$$\Phi_t + H = \text{const}$$

where

$$H = \frac{c^2}{2} + \frac{p}{\rho} = \frac{u^2 + v^2}{2} + \frac{p}{\rho}$$

For small perturbations from the steady flow

$$\varphi_t + c\delta c + \frac{\delta p}{\rho} = 0 \ldots \ldots \ldots \ldots [2]$$

Equations [1] and [2] are valid in the region $-\infty < x < 0$.

Entry to the Cascade. The distance 1–3 is assumed small so that unsteady inertia effects between 1 and 3 may be neglected. Thus

$$\left.\begin{array}{l} c_1^2/2 + p_1/\rho = c_3^2/2 + p^3/\rho \\ c_1\delta c_1 + \delta p_1/\rho = c_3\delta c_3 + \delta p_3/\rho \end{array}\right\} \quad \ldots \ldots \ldots [3]$$

At 1, Equation [2] may be rewritten

$$(\varphi_t)_1 + c_3\delta c_3 + \delta p_3/\rho = 0 \ldots \ldots \ldots \ldots [4]$$

Momentum Effects in the Cascade. Considering the flow between 3 and 2 as one dimensional, the momentum equation in the l-direction is

$$\partial c/\partial t + c\partial c/\partial l + 1/\rho \partial p/\partial l = 0$$

Integrating with respect to l from 3 to 2, with the assumption that $c = c_3$ between 3 and 2 and the change in velocity from c_3 to c_2 takes place in a short distance

$$L\partial c_2/\partial t + c_2^2/2 - c_3^2/2 + p_2/\rho - p_3/\rho = 0$$

For small perturbations from a steady flow

$$L\partial(\delta c_3)/\partial t + c_2\delta c_2 - c_3\delta c_3 - \delta p_3/\rho = 0 \ldots \ldots [5]$$

if p_2 does not change. Eliminating $(c_3\delta c_3 + \delta p_3/\rho)$ between 4 and 5

$$L\frac{\partial(\delta c_2)}{\partial t} + (\varphi_t)_1 + c_2\delta c_2 = 0 \ldots \ldots \ldots \ldots [6]$$

From continuity

$$A_2 c_2 = A_3 c_3$$
$$A_2\delta c_3 = A_3\delta c_2 + c_3\delta A_2$$

Thus

$$\delta c_2 = \frac{A_3}{A_2}\delta c_3 - c_2\frac{\delta A_2}{A_2}$$

and

$$c_2\delta c_2 = \left(\frac{A_3}{A_2}\right)^2 c_3\delta c_3 - \left(\frac{A_3}{A_2}\right)^2 c_2^2\frac{\delta A_2}{A_2}$$

Also from continuity $u = c_2 \cos \beta_2$ so that $\delta u = \varphi_x = \delta c_2 \cos \beta_2$. Substitution for c_2, δc_2 yields

$$c_2 \delta c_2 = \left(\frac{A_1}{A_2}\right)^2 \frac{c_x \varphi_x}{\cos^2 \beta_2} - \left(\frac{A_1}{A_2}\right)^2 \frac{c_x{}^2}{\cos^2 \beta_2} \frac{\delta A_2}{A_2}$$

By substituting this expression for $c_2 \delta c_2$ in Equation [6] we obtain, at point 1, the following equation

$$\varphi_t + \frac{L}{\cos \beta_2} \varphi_{xt} + \frac{c_x \varphi_x}{\alpha^2 \cos^2 \beta_2} - \frac{c_x{}^2}{\alpha^2 \cos^2 \beta_2} \frac{\delta A_2}{A_2} = 0 \ldots [7]$$

But

$$\frac{\delta A_2}{A_2} = \frac{\delta \alpha}{\alpha} = \frac{1}{\alpha} \frac{d\alpha}{d(\cot \beta_1)} \delta(\cot \beta_1)$$

$$= \frac{\alpha'}{\alpha} \delta\left(\frac{u}{v}\right) \text{ where } \alpha' = \frac{d\alpha}{d(\cot \beta_1)}$$

$$= \frac{\alpha'}{\alpha} \frac{\varphi_x - \cot \beta_1 \varphi_y}{c_y}$$

Thus, at 1

$$\varphi_t + \frac{L}{\cos \beta_2} \varphi_{xt} + \frac{c_x \varphi_x}{\alpha^2 \cos^2 \beta_2}\left[1 - \frac{\cot \beta_1 \alpha'}{\alpha}\right]$$
$$+ \frac{\cot^3 \beta_1 \alpha'}{\alpha^3 \cos^3 \beta_2} c_x \varphi_y = 0 \ldots\ldots\ldots\ldots\ldots [8]$$

This equation must be satisfied by φ at the entrance to the cascade; that is, at $x = 0$. Substituting the solution for φ into Equation [8] and separating coefficients of $\cos(n\pi y/b)$, $\sin(n\pi y/b)$ two simultaneous differential equations are obtained in $a_n(t)$ and $b_n(t)$

$$a_n\left[\left(\cos \beta_2 + \frac{Ln\pi}{b}\right)D + \frac{n\pi}{b}\frac{c_x}{\alpha^2 \cos \beta_2}\left(1 - \frac{\cot \beta_1 \alpha'}{\alpha}\right)\right]$$
$$+ b_n\left[\frac{n\pi}{b}\frac{\cot^2 \beta_1 \alpha' c_x}{\alpha^3 \cos \beta_2}\right] = 0$$

$$b_n\left[\left(\cos \beta_2 + \frac{Ln\pi}{b}\right)D + \frac{n\pi}{b}\frac{c_x}{\alpha^2 \cos \beta_2}\left(1 - \frac{\cot \beta_1 \alpha'}{\alpha}\right)\right]$$
$$- a_n\left[\frac{n\pi}{b}\frac{\cot^2 \beta_1 \alpha' c_x}{\alpha^3 \cos \beta_2}\right] = 0$$

Assuming solutions of the form

$$a_n = A_n e^{\lambda_n t}$$
$$b_n = B_n e^{\lambda_n t}$$

the characteristic equation

$$\left[\left(\cos \beta_2 + \frac{Ln\pi}{b}\right)\lambda_n + \frac{n\pi}{b}\frac{c_x}{\alpha^2 \cos \beta_2}\left(1 - \frac{\cot \beta_1 \alpha'}{\alpha}\right)\right]^2$$
$$+ \left[\frac{n\pi}{b}\frac{\cot^2 \beta_1 \alpha' c_x}{\alpha^3 \cos \beta_2}\right]^2 = 0 \ldots\ldots [9]$$

is obtained.

The roots of this equation may represent oscillatory disturbances which are damped out, amplified, or persist unchanged depending on the value of the damping term containing $[1 - (\cot \beta_1 \alpha'/\alpha)]$. If $\alpha' < \alpha/(\cot \beta_1)$, the disturbance will die away. If $\alpha' > \alpha/(\cot \beta_1)$, it will be amplified beyond the range in which a linearized analysis is valid. If $\alpha' = \alpha/(\cot \beta_1)$ (Fig. 5), the disturbance will persist and a_n, b_n will be of the form $A_n \cos \omega_n t +$
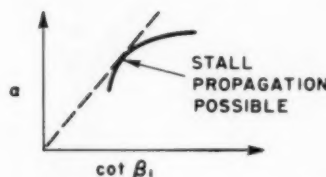
FIG. 5 STABILITY CRITERION FOR STALL PROPAGATION

$B_n \sin \omega_n t$, $A_n \sin \omega_n t - B_n \cos \omega_n t$, respectively. The potential function then represents waves traveling along the cascade. At this point, the characteristic equation becomes

$$\lambda_n{}^2 + \left[\frac{\frac{n\pi}{b} \cot \beta_1 c_x}{\alpha^2 \cos \beta_2 \left(\frac{Ln\pi}{b} + \cos \beta_2\right)}\right]^2 = 0$$

representing waves of radian frequency

$$\omega_n = \frac{\frac{n\pi}{b} \cot \beta_1 c_x}{\alpha^2 \cos \beta_2 \left[\frac{Ln\pi}{b} + \cos \beta_2\right]}$$

The velocity of wave propagation of the nth harmonic along the cascade V_n = frequency × wave length

$$V_{pn} = \frac{\omega_n}{2\pi} \frac{2b}{n}$$

$$= \frac{\cot \beta_1 c_x}{\alpha^2 \cos \beta_2 \left[\frac{Ln\pi}{b} + \cos \beta_2\right]}$$

$$\frac{V_{pn}}{c_x} = \frac{\cot \beta_1}{\alpha^2 \cos \beta_2 \left[\frac{Ln\pi}{b} + \cos \beta_2\right]} \ldots\ldots\ldots [10]$$

This result indicates that the velocity of stall propagation increases with the size of the stall cell, the limiting value of V_p/c_x for stall cells covering many blades being $(\cot \beta_1)/(\alpha^2 \cos^2 \beta_2)$. In Figs. 6 and 7 $(V_p \alpha^2)/c_x$ is plotted against β_1 for cascades with 10 deg and 15 deg turning and different values of $\pi L/b$. The velocity of propagation is taken as the velocity of the fundamental component of the wave. The expression for V_{pn}/c_x includes the number of the harmonic n and indicates that higher harmonics travel more slowly than the primary wave. This is not in accord with experience and shows that a linearized analysis is inadequate in this respect.

The propagation velocity may be related conveniently to the cascade pressure rise (which is a more familiar parameter than α) by replacing α with its equivalent

$$\frac{\cos \beta_1}{\cos \beta_2 \sqrt{(1 - C_p)}}$$

Then we obtain

$$\frac{V_{pn}}{c_x} = \frac{\cot \beta_1(1 - C_p)}{\cos^2 \beta_1 \left(\frac{Ln\pi}{b \cos \beta_2} + 1\right)}$$

that is
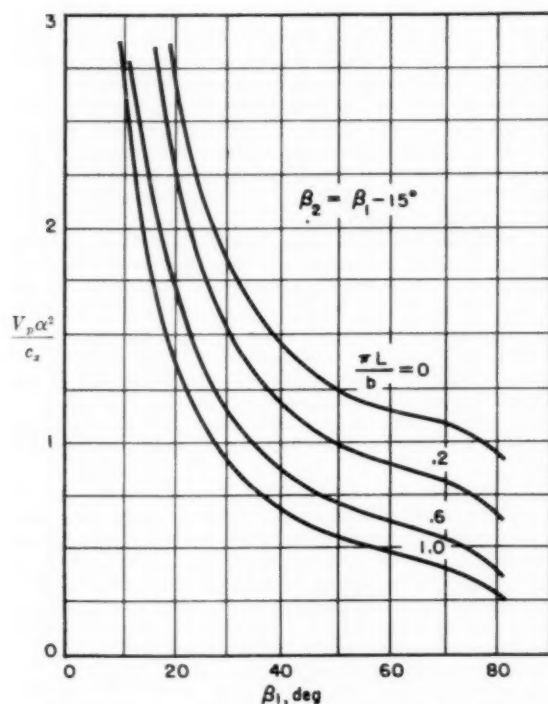
FIG. 6 PREDICTED PROPAGATION VELOCITY FOR CASCADES WITH
15-DEG TURN
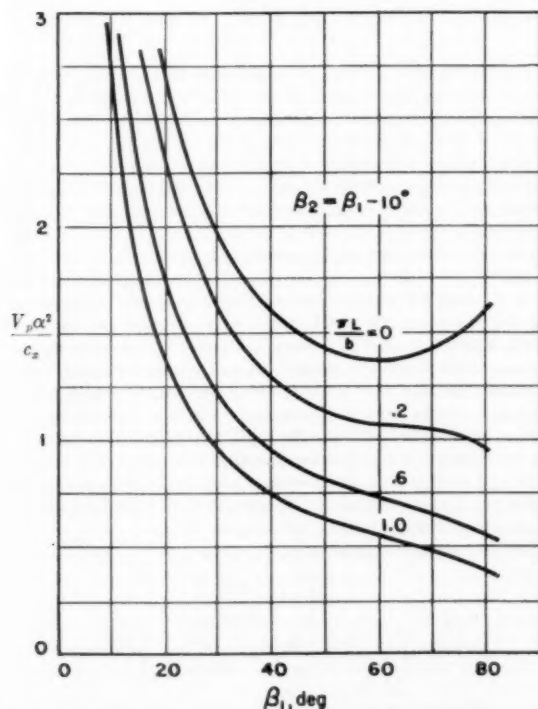
FIG. 7 PREDICTED PROPAGATION VELOCITY FOR CASCADES WITH
10-DEG TURN

$$\frac{V_{pn}}{c_s} = \frac{2(1 - C_p)}{\sin 2\beta_1 \left(\dfrac{Ln\pi}{b \cos \beta_2} + 1 \right)} \qquad [11]$$

For stall cells covering many blades, this expression reduces to

$$\frac{V_p}{c_s} = \frac{2(1 - C_p)}{\sin 2\beta_1}$$

It should be noted that, with the boundary-layer time delay neglected, no limitation is placed on the wave length of the disturbances. Consideration of the boundary-layer delay permits prediction of the size of the wave in addition to the velocity.

2.4 *Solution for $\tau = 0$ With Downstream Mixing.* By assuming that the fluid from each blade passage mixes without pressure recovery or angle change just before leaving the cascade so that the downstream flow field is a continuum, it is possible to obtain another expression for propagation velocity. An analysis of this case (4) yields the same stability criterion for stall propagation as in the previous section and the equation for propagation velocity

$$\frac{V_p}{c_s} = \frac{2(1 - C_p)}{\sin 2\beta_1 \left[\dfrac{L\pi}{b \cos \beta_2} + 2 \right]} \qquad [12]$$

For very large wave lengths, the propagation velocity is half that given by Equation [11].

2.5 *Solution Including Boundary-Layer Delay With No Downstream Mixing.* The time constant τ of the boundary-layer delay will now be included in the analysis to determine the additional limitations it imposes on stall propagation. Returning to Equation [7]

$$\varphi_t + \frac{L}{\cos \beta_2} \varphi_{st} + \frac{c_s \varphi_s}{\alpha^2 \cos^2 \beta_1} - \frac{c_s^2}{\alpha^3 \cos^2 \beta_1} \frac{\delta \alpha}{\alpha} = 0$$

and setting

$$\delta \alpha = \frac{\alpha'}{\tau D + 1} \delta(\cot \beta_1)$$

the equation satisfied by φ at 1 is

$$\tau \varphi_{tt} + \frac{L\tau}{\cos \beta_2} \varphi_{stt} + \varphi_t + \varphi_{st}\left[\frac{L}{\cos \beta_2} + \frac{\tau c_s}{\alpha^2 \cos^2 \beta_2} \right]$$
$$+ \frac{c_s \varphi_s}{\alpha^2 \cos^2 \beta_2}\left[1 - \frac{\cot \beta_1 \alpha'}{\alpha} \right] - \frac{\cot^2 \beta_1 \alpha'}{\alpha^3 \cos^2 \beta_2} c_s \varphi_y = 0 \quad [13]$$

Inserting the solution for φ of Equation [1] and separating the coefficients of $\cos n\pi y/b$, $\sin n\pi y/b$ gives two second-order differential equations in a_n, b_n. Taking solutions of the form

$$a_n = A_n e^{\lambda_n t}$$
$$b_n = B_n e^{\lambda_n t}$$

a characteristic equation for λ_n is found

$$\left\{ \left[\tau + \frac{n\pi}{b} \frac{L\tau}{\cos \beta_2} \right] \lambda_n^2 + \left[1 + \frac{n\pi}{b} \left(\frac{L}{\cos \beta_2} + \frac{\tau c_s}{\alpha^2 \cos^2 \beta_2} \right) \right] \lambda_n \right.$$
$$\left. + \frac{n\pi}{b} \frac{c_s}{\alpha^2 \cos^2 \beta_2}\left[1 - \frac{\cot \beta_1 \alpha'}{\alpha} \right] \right\}^2$$
$$+ \left(\frac{n\pi}{b} \frac{\cot^2 \beta_1 \alpha' c_s}{\alpha^3 \cos^2 \beta_2} \right)^2 = 0 \quad [14]$$

This equation will be examined to find the conditions for undamped oscillations in time.

Equation [14] is of form

$$[d_1\lambda_n{}^2 + d_2\lambda_n + d_3]^2 + d_4{}^2 = 0 \dots\dots\dots [15]$$

For undamped oscillations, $\lambda_n = \pm i\omega_n$ must be a root. With $\lambda_n = +i\omega_n$ a solution

$$[-d_1\omega_n{}^2 + id_2\omega_n + d_3]^2 + d_4{}^2 = 0 \dots\dots\dots [16]$$

With $\lambda_n = -i\omega_n$ a solution

$$[-d_1\omega_n{}^2 - id_2\omega_n + d_3]^2 + d_4{}^2 = 0 \dots\dots\dots [17]$$

Subtracting Equation [17] from [16] and factoring

$$2[-d_1\omega_n{}^2 + d_3][2id_2\omega_n] = 0$$

Thus if $\lambda_n = \pm i\omega_n$ is a root of Equation [15] then $\omega_n{}^2 = d_3/d_1$. Putting $\omega_n = \sqrt{(d_3/d_1)}$ in Equation [15] then

$$\left[-d_3 + id_2 \sqrt{\left(\frac{d_3}{d_1}\right)} + d_3 \right]^2 + d_4{}^2 = 0$$

$$-d_2{}^2 \frac{d_3}{d_1} + d_4{}^2 = 0$$

or

$$d_3/d_1 = (d_4/d_2)^2 \dots\dots\dots\dots\dots [18]$$

The requirements for an undamped oscillatory solution to Equation [15] are, therefore, that $d_3/d_1 = (d_4/d_2)^2$. With this requirement satisfied, two of the roots of the equation are $\pm i\omega_n$ where $\omega_n = d_4/d_2$.

A factor of the equation is then $[\lambda_n{}^2 + (d_4/d_2)^2]$ and the remaining two roots satisfy the equation

$$\lambda_n{}^2 + 2\left(\frac{d_2}{d_1}\right)\lambda_n + \left[\left(\frac{d_4}{d_2}\right)^2 + \left(\frac{d_2}{d_1}\right)^2\right] = 0 \dots\dots [19]$$

For the remaining two roots to be stable the coefficient of λ_n in Equation [19] must be positive; i.e., d_2/d_1 must be positive.

It can be seen, therefore, that for Equation [15] to have roots representing undamped oscillations of frequency ω_n:

(a) All coefficients must be positive
(b) $d_3/d_1 = (d_4/d_2)^2$

Under these circumstances, $\omega_n = d_4/d_2$.

Applying these results to Equation [14], the requirement for stall propagation is

$$\frac{\left[\dfrac{n\pi}{b} \dfrac{c_x}{\alpha^2 \cos^2\beta_2}\left(1 - \dfrac{\cot\beta_1\alpha'}{\alpha}\right)\right]}{\tau\left[1 + \dfrac{n\pi}{b}\dfrac{L}{\cos\beta_2}\right]}$$

$$= \frac{\left[\dfrac{n\pi}{b}\dfrac{\cot^2\beta_1\alpha'c_x}{\alpha^2\cos^2\beta_2}\right]^2}{\left[1 + \dfrac{n\pi}{b}\left(\dfrac{L}{\cos\beta_2} + \dfrac{\tau c_x}{\alpha^2\cos^2\beta_2}\right)\right]^2} \dots [20]$$

and the frequency of the oscillations at any point in radians/second is

$$\omega_s = \frac{\left[\dfrac{n\pi}{b}\dfrac{\cot^2\beta_1\,\alpha'c_x}{\alpha^2\cos^2\beta_2}\right]}{1 + \dfrac{n\pi}{b}\left(\dfrac{L}{\cos\beta_2} + \dfrac{\tau c_x}{\alpha^2\cos^2\beta_2}\right)} \dots\dots [21]$$

If $\tau = 0$ the same results are obtained as in the earlier analysis.

The importance of this solution lies, not in the precise relationship, but in the fact that a relation is established between L/b, τ, α, α', β_1, and β_2 for stall propagation to be possible. More explicitly, since for a given cascade τ, α, α', and β_2 all depend on β_1, a relationship has been found between L/b and β_1 for stall propagation. Thus, over a range of inlet angles, stall propagation can occur and the size of the stall cell will depend on the inlet angle. This result may help to explain the variation in size and number of stall cells found in an axial-compressor stage over the working range of angle of attack.

With the boundary-layer delay included in the analysis, stall propagation occurs with $(\cot\beta_1\alpha'/\alpha)$ less than unity, so that the frequency of the disturbances is lower than the frequency of the stalls of the same wave length when the boundary layer is left out. The boundary-layer delay therefore has the effect of decreasing the velocities of propagation of the stalled regions. The boundary-layer time constant should be approximately of the order L/c_3 so that the boundary-layer delay may be of equal importance with the inertia delay of the fluid within the cascade and will have little effect on the velocity of propagation of large stall cells.

Calculation of Wave Length When Rotating Stall Commences. It is now possible, without making any additional assumptions, to predict the wave length of the disturbances when rotating stall commences. Because of the simple representation employed for the boundary-layer response it would be unwise to attempt to derive numerical values from this result, but it is of interest inasmuch as the main features of the phenomenon can be reproduced with a simple model.

Returning to Equation [20], which can be rewritten

$$\frac{\left[1 - \dfrac{\cot\beta_1\alpha'}{\alpha}\right]}{\left[\dfrac{\tau\cot^4\beta_1\alpha'^2c_x}{\alpha^4\cos^2\beta_2}\right]} = \frac{\dfrac{n\pi}{b}\left[1 + \dfrac{n\pi}{b}\dfrac{L}{\cos\beta_2}\right]}{\left[1 + \dfrac{n\pi}{b}\left(\dfrac{L}{\cos\beta_2} + \dfrac{\tau c_x}{\alpha^2\cos^2\beta_2}\right)\right]^2}$$

as the requirement for stall propagation to be possible, it can be seen that as the inlet angle β_1 to the cascade is increased and flow separation commences on the airfoils, α' will increase from almost zero to a positive quantity. The group on the left side of the equation (denoted below by A) will therefore have a value which begins as a large positive quantity at the design point and decreases as β_1 is increased. At any setting of β_1, the group on the right side of the equation (denoted by C) can have an infinite number of values for each harmonic, depending on the value of b assumed. Considering only the primary component of the wave $(n = 1)$, there will however be a maximum value of C corresponding to a wave length which might be considered as the wave length which is closest to being propagated. As β_1 is increased, the maximum possible value of C can be computed for each value of b and compared with the value of A (see Fig. 8), which will in general be larger than C. However, as β_1 is increased the difference between A and C shrinks until, at the critical value of β_1, the maximum value of C corresponds to the value of A at that point and stall propagation becomes possible for the wave length which yields this value of C. The problem is, therefore, the determination of the value of b which gives $(C)_{max}$.

For the primary component of the wave

$$C = \frac{\pi\left[b + \dfrac{\pi L}{\cos\beta_2}\right]}{\left[b + \pi\left(\dfrac{L}{\cos\beta_2} + \dfrac{\tau c_x}{\alpha^2\cos^2\beta_2}\right)\right]^2}$$

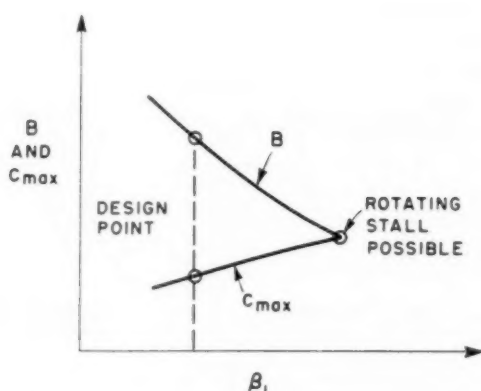For a maximum (or minimum) value of C, $(dC)/(db) = 0$, and this occurs when

FIG. 8 STABILITY CRITERION WITH BOUNDARY-LAYER DELAY

$$ b = \frac{\pi L}{\cos \beta_2} \left[\frac{\tau}{\alpha^2 (L/c_3)} - 1 \right] $$

A study of the expression for C shows that this value of b gives a maximum value of C, at a value of b which may be positive or negative depending on whether $\frac{\tau}{\alpha^2(L/c_3)}$ is greater than, or less than unity.

There are now two possibilities. If $\frac{\tau}{\alpha^2(L/c^3)} > 1$, then for positive (i.e., real) values of b, C has a maximum when $b = \frac{\pi L}{\cos \beta_2} \left[\frac{\pi}{\alpha^2(L/c_3)} - 1 \right]$ and this will be the value of b when stall propagation commences. If $\frac{\tau}{\alpha^2(L/c_3)} < 1$, then the maximum value of C for positive values of b occurs when $b = 0$, and stall propagation will commence with the smallest values of b that are physically possible. Thus, the wave length is governed by the ratio $\tau/(L/c_4)$ which is the ratio of the boundary-layer time delay to the inertia delay of the fluid within the cascade.

2.6 *Discussion of the Analysis.* Solutions may be obtained for cascades with inlet guide vanes (4) and for the case with both the boundary-layer delay and downstream mixing included, but the principal results are now clear. With the boundary-layer delay left out, stall propagation becomes possible for all wave lengths at the same time. With the boundary-layer delay included, stall propagation commences with a wave length which depends on the ratio of the boundary-layer delay to the inertia delay of the fluid within the passage. The assumption of immediate mixing after the cascade leads to a lower propagation velocity than does the assumption of no mixing and constant downstream pressure.

The solutions may be checked against the earlier analyses of Emmons (1), Sears (5), and Marble (6). Emmons' analysis corresponds to the case with no downstream pressure fluctuations, finite boundary-layer delay, and the wave length b and blade chord L becoming vanishingly small. For this case Equations [20] and [21] yield $\alpha' = \alpha/\cot \beta_1$ as the stability criterion and $\omega_a = \cot \beta_1/\tau$, in agreement with Emmons' analysis. In Sears' "channel" theory with the phase lag set equal to zero, and in Marble's theory, the basic assumptions are that the pressure rise across the cascade is small, blade-chord effects are neglected, and the downstream field is considered to be continuous. For this case Equation [11] applies, and the expression for propagation velocity obtained from the present analysis is identical with Sears' and Marble's results; namely $V_p/c_x = 1/\sin 2\beta_1$.

Thus, the present theory is consistent with the earlier analyses

when the same assumptions are employed. Sears' "airfoil" theory yields results which are radically different from the other theories because it is based on a different cascade representation which neglects losses.

3 EXPERIMENTAL INVESTIGATION OF ROTATING STALL

Having established solutions for the two limiting downstream conditions, it is now desirable to determine experimentally the true conditions after the cascade, and to compare the predicted propagation velocity with the experimentally observed values.

Rotating stall has been studied (4) in the circular cascade shown diagrammatically in Fig. 9. Air flows outward through a set of variable-angle nozzles and enters the cascade at an angle of attack which may be controlled by changing the nozzle angle. At high angles of attack, rotating stall occurs and may be observed using Schlieren techniques through windows in the side of the test section. In Fig. 10, α is plotted against $\cot \beta_1$ under steady flow conditions, the left-hand end of the curve corresponding to the point where rotating stall commenced. Just before rotating stall begins, the slope of the curve is close to the value predicted by the simple theory of sections 2.3 and 2.4. The number of stall cells varied from 9 to 12, and the downstream pressure fluctuations were of the order of 30 per cent of the upstream pressure fluctuations. Using the simple theory of section 2.2, neglecting downstream pressure fluctuations and boundary-layer delay, the theoretical and observed propagation velocities are shown in Fig. 11. Since the desired cascade performance in terms of C_p and β_2 as functions of β_1 can be measured only in steady flow, it is necessary to extrapolate the curves into the rotating stall region to obtain values of theoretical propagation velocity. This was done by assuming that, in the absence of rotating stall, $(dC_p)/(d\beta_1)$ and $(d\beta_2)/(\beta_1)$ would have the same values as those just before rotating stall began. The theoretical velocities are slightly higher than the experimental values. If the theory of section 2.4 is employed, the predicted velocities are much lower than the observed values. As expected, the experimental values are bracketed by the solutions for the two limiting downstream conditions, and in this case neglecting downstream pressure fluctuations is the better approximation of the two. The experimental propagation velocities in the cascade do not show any consistent effect of stall cell number.

From the high-speed Schlieren photographs, the magnitude of the boundary-layer delay may be estimated. For example, in Fig. 12, the stalling process on the center airfoil covers 2–3 frames from the first signs of separation to complete stall. Since the film speed is 5000 frames per sec, the time for stall is approximately 0.5 millisec. This is not equal to the boundary-layer time delay τ, since for the equivalent exponential lag system the time required to attain 95 per cent of the final change in α after a step change in β_1 would be 3τ. An approximate value for τ is therefore 0.16 millisec. The inertia time constant L/c_3 was equal to 0.35 millisec since c_3 was 230 fps and the blade chord L was 0.96 in. Thus, the upper limit to the value of τ was approximately $1/2(L/c_3)$ and the true value was certainly smaller than this since the inlet angle did not change instantaneously, and the observed response of the boundary layer was therefore slower than the theoretical response to a step change in inlet angle. Using this value of τ in the expression for the wave length when rotating stall commences, $\left[\frac{\tau}{\alpha^2(L/c_3)} - 1 \right]$ has a value of -0.1, suggesting that rotating stall should commence with the largest number of stall cells that are physically possible. While this conclusion is a reasonable agreement with the facts for the circular cascade, since the number of stall cells is quite large, other observers have frequently found that rotating stall commenced with only one
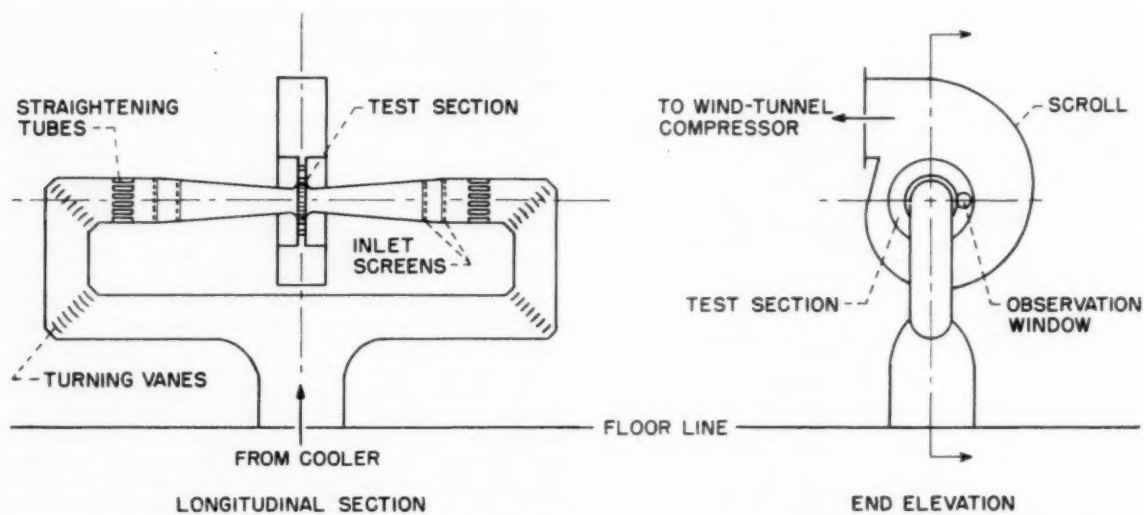
FIG. 9   SCHEMATIC DIAGRAM OF CIRCULAR CASCADE
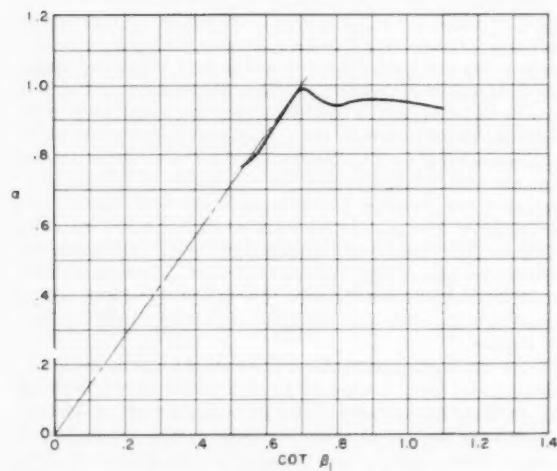
FIG. 10   CASCADE PERFORMANCE α VERSUS COT β_1
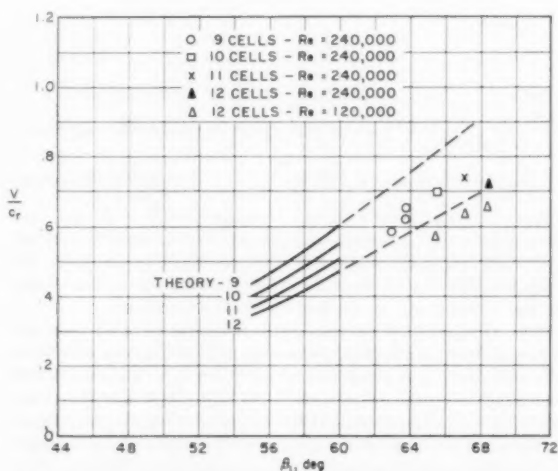
FIG. 11   THEORETICAL AND EXPERIMENTAL VALUES OF PROPAGATION VELOCITY
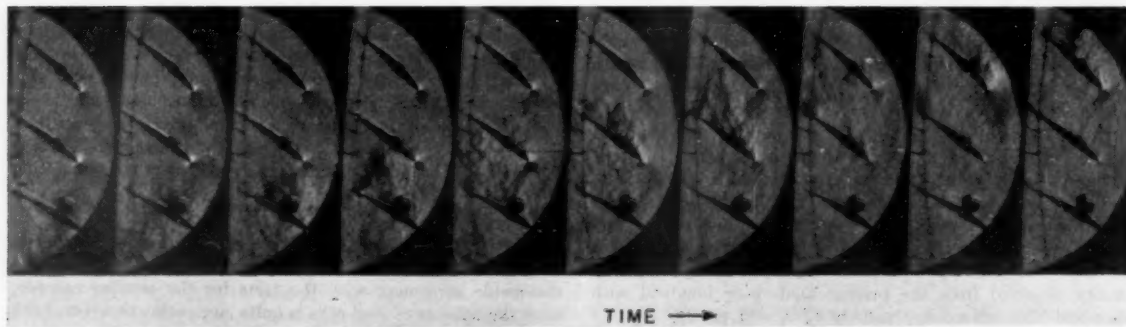
FIG. 12   SCHLIEREN VIEWS OF ROTATING STALL—5000 FRAMES PER SEC
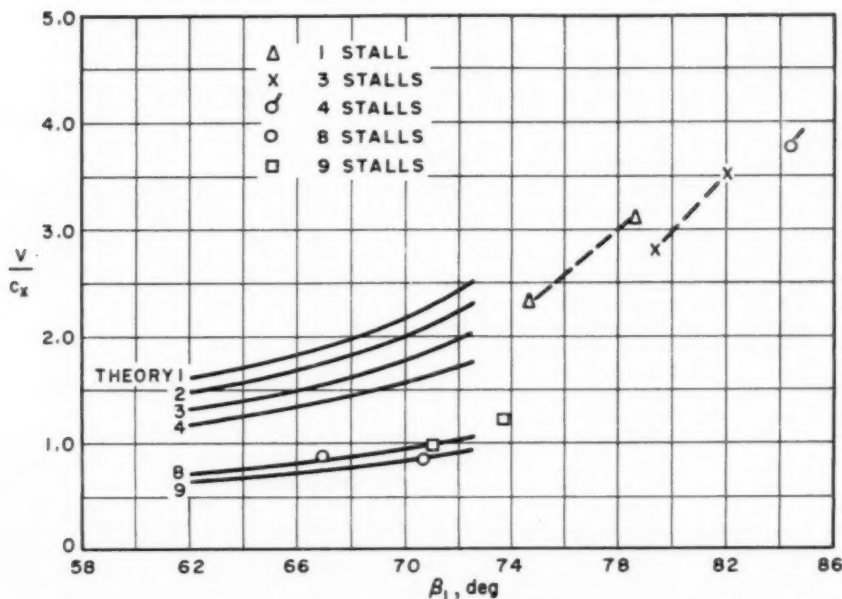
FIG. 13   THEORETICAL AND EXPERIMENTAL PROPAGATION VELOCITIES FOR A ROTOR WITH GUIDE VANES
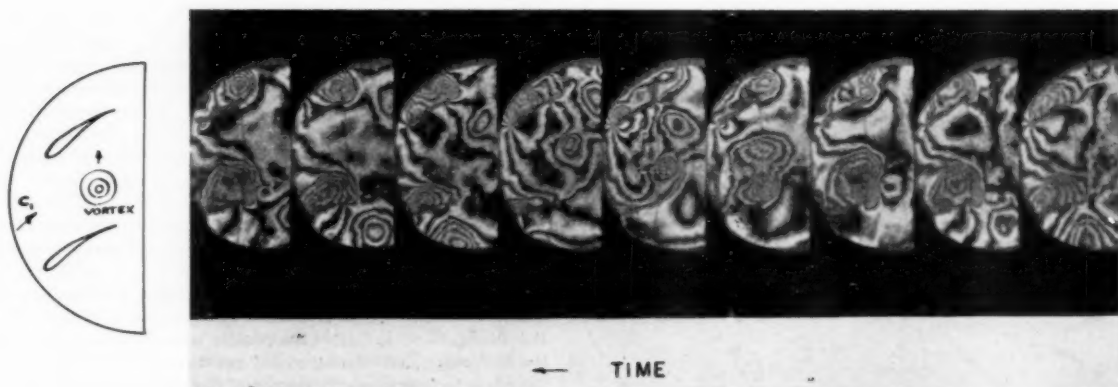
FIG. 14   INTERFEROGRAMS OF ROTATING STALL—5000 FRAMES PER SEC

stall cell, and the present theory is unable to explain this because in any practical machine the value of $\left[\dfrac{\tau}{\alpha^2(L/c_3)} - 1\right]$ will always be close to zero. Prof. R. E. Kronauer of Harvard has pointed out that if α is assumed to be a function of $\partial\beta/\partial x$ and $\partial\beta/\partial y$ as well as of β, the resulting stability requirements permit a much greater variety of wave lengths than does the present analysis.

An experimental study of rotating stall in an axial compressor consisting of a set of inlet guide vanes and a rotor only (4) has yielded results similar to those found in the cascade. The downstream pressure fluctuations were approximately 25 per cent of the upstream fluctuations, and the simple theory of section 2.3 predicted the propagation velocity quite well, Fig. 13, while the application of the theory of section 2.4 resulted in values of propagation velocity which were much lower than the observed values.

In both the cascade and the compressor the velocity of the air within the stall cells was close to zero, so that the linearized analy-sis could not be expected to give accurate prediction- of propagation velocity. In addition, the appearance of low-tensity spots within the regions of flow separation in many of the schlieren photographs suggested that the circulation was shed in a chunk as a vortex. To check this hypothesis, interferograms were taken of the unsteady flow in the circular cascade, and the presence of these vortexes was immediately confirmed. In Fig. 14 it can be seen that the airfoil in the observation window sheds a large vortex as the circulation shed earlier from the preceding blade passes close to the trailing edge. The shed vortex drifts towards the next airfoil and is followed by a starting vortex as the normal circulation on the blade is restored. The lines in the photograph are contours of constant density. Hot-wire measurements in the compressor indicated that the shed vortexes decayed quite slowly after leaving the stalled blade row, and that for a considerable distance downstream the velocity profile approximated the modified square wave produced by a vortex street.

4 ANALYSIS OF A VORTEX MODEL OF STALL PROPAGATION

4.1 Effect of a Vortex on an Airfoil. It is seen in Fig. 14 that an airfoil stalls by shedding its lift in the form of a discrete vortex which accumulates at and departs from its leading edge when a vortex from the previous blade passes its trailing edge. It is therefore of interest to investigate the effect of a vortex on an airfoil in potential flow. Considering the simplest possible case of a vortex and a flat plate satisfying the Joukowski trailing-edge condition in steady flow, Fig. 15 shows loci of the vortex position for constant induced plate circulation. Γ_0 is the induced plate circulation, Γ the circulation of the vortex, and r the ratio Γ_0/Γ. As a vortex approaches the plate from below, at the same time passing from left to right, the plate circulation remains small or negative until the vortex approaches the trailing edge, then increases rapidly. For a real fluid, stalling of the plate will occur at some limiting value of r. Consideration of the presence of vorticity shed by the plate as its circulation changes in the unsteady flow does not seriously modify these conclusions. Numerical calculations of the trajectory of a vortex shed from an airfoil in a cascade (8) show that the vortex moves along a path similar to that found experimentally from the interferograms. Unfortunately, due to the complexity of the problem, it has not yet been found possible to represent the shedding process and the accompanying flow changes simply enough to yield a completely analytical solution to the problem. However, a solution can be obtained from a study of the downstream field, using experimental data to represent the dynamic characteristics of the cascade.
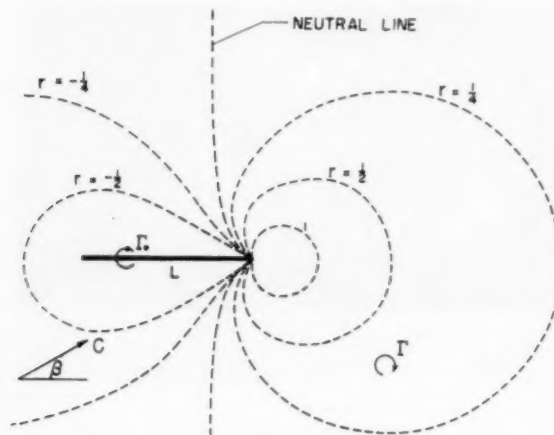
FIG. 15 CIRCULATION INDUCED ON A PLATE BY A VORTEX

4.2 Derivation of Expressions for Propagation Velocity Velocity measurements taken during stall propagation indicate that for a given operating condition or mean flow, the unsteady velocity perturbation translates at constant velocity V_p along the cascade, and to a first approximation, the shape of the velocity profile of the unsteady perturbation is maintained constant. With the approximation that each cascade airfoil sheds a discrete potential vortex $(+\Gamma)$ when it enters the stall cell and another $(-\Gamma)$ of opposite sign when it leaves the cell, it follows that all the trajectories of vortexes (of same sign) shed from the cascade airfoils must be translates of the same curve and separated by the blade spacing(s) along the cascade.

The vortexes which have been shed downstream must be arranged in a pattern as shown in Fig. 16. The dashed lines indicate the trajectories of the vortexes shed from two succeeding
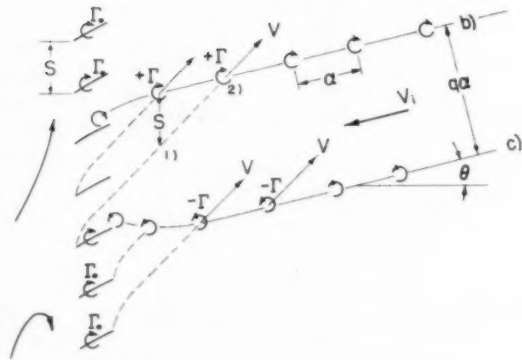
FIG. 16 PATTERN OF VORTEXES IN A STALL CELL

blades. The $(+\Gamma)$ vortexes must lie on a line (b) which eventually becomes straight downstream of the cascade and the $(-\Gamma)$ vortexes must be located along a line (c) which becomes straight and parallel to line (b) downstream of the cascade. The region between lines (b) and (c) is called the stall cell. Lines (b) and (c) are of constant shape in time and move relative to the cascade with the velocity of propagation, V_p. It might be expected that the vortex street of Fig. 16 should have the classic proportions required by von Karman's stability analysis (9). However, the measured width of the street appeared to be an order of magnitude greater than the stability analysis predicts.

In a co-ordinate system which is fixed to the stall cell, the blades move down in Fig. 16 with velocity V_p; and the shed vortexes move downstream along the stationary path lines (b) and (c). The flow is steady in time except for the effects due to the finite spacing of blades and shed vortexes. It can be seen from Equations [22], which follow, that these are local effects confined to an area within approximately one blade spacing of the cascade and one vortex spacing of lines (b) and (c). Flow is diverted to each side of the stall cell which has the nature of a thick wake.

After they are far downstream of the cascade, the shed vortexes are spaced uniformly by a distance (a) along lines (b) and (c), and move with transport velocity V. θ is defined as the angle between each street and the cascade axis, and q is defined as the number of vortex spaces between the streets as indicated in Fig. 16. In Fig. 17 are indicated the velocity triangles associated with the stall cell. The velocity at any point far downstream of the cascade is C_2 plus the induced velocity due to the stall cell (which is the velocity induced by two infinite vortex streets).

If the blades stall out completely when they enter the cell, the strength of the shed vortexes will be of the order of magnitude

$$\pm\Gamma = (C_{y_1} - C_{y_2})s$$

From (9), for a single infinite row of equidistant vortexes, each of strength Γ (+ clockwise) at distances a apart, with the origin at a vortex, and the axis of x along the row, the velocity components induced by the vortexes are
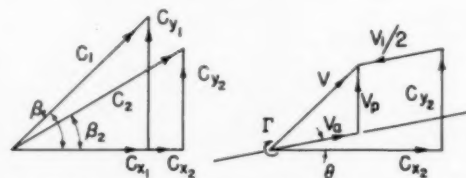
FIG. 17 VELOCITY TRIANGLES FOR STALL CELL

$$u = \frac{\Gamma}{2a} \frac{\sinh{(2\pi y/a)}}{\cosh{(2\pi y/a)} - \cos{(2\pi x/a)}} \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} \dots [22]$$

$$v = \frac{\Gamma}{2a} \frac{\sin{(2\pi x/a)}}{\cosh{(2\pi y/a)} - \cos{(2\pi x/a)}}$$

Therefore

$$\frac{\partial v}{\partial x} = \frac{-\dfrac{\Gamma}{2a}\dfrac{2\pi}{a}\left(\cosh\dfrac{2\pi y}{a} - \cos\dfrac{2\pi x}{a}\right)\cos\dfrac{2\pi x}{a} - \sin^2\dfrac{2\pi x}{a}}{(\cosh 2\pi\, y/a) - (\cos 2\pi x/a)^2}$$

and for v_{max}

$$\cosh\frac{2\pi y}{a}\cos\frac{2\pi x}{a} - 1 = 0$$

$$\cos\frac{2\pi x}{a} = \frac{1}{\cosh\dfrac{2\pi y}{a}}$$

$$\sin\frac{2\pi x}{a} = \left(1 - \cos^2\frac{2\pi x}{a}\right)^{1/2} = \frac{\left(\cosh^2\dfrac{2\pi y}{a} - 1\right)^{1/2}}{\cosh\dfrac{2\pi y}{a}}$$

$$= \tanh\frac{2\pi y}{a}$$

$$v_{max} = \frac{\Gamma}{2a\sinh 2\pi y/a}$$

Therefore, far downstream, the maximum lateral velocity (normal to lines b and c) the vortex streets can induce on each other is

$$v_{max} = \frac{\Gamma}{2a\sinh 2\pi q}$$

where qa is the distance between lines (b) and (c). If $q > 1$

$$v_{max} \leq \frac{C_{y1} - C_{y2}}{500}\frac{s}{a}$$

and may be neglected.

The basis for taking $q > 1$ is derived from the hot-wire data taken downstream of the isolated rotor where it is noted that the cells always cover at least two blade spaces. This appears to be true in general for stall cells in rotors. In the interferometer picture from the circular cascade, it is noted that $q \approx 1$; however, the downstream flow field extends only two or three blade chords from the circular cascade, so the initial assumption of an infinite downstream field is not valid for the circular cascade.

More than a distance (a) from the vortex streets, the longitudinal velocity induced by them may be neglected outside the cell, while inside the cell, from Equation [22]

$$V_i = \Gamma/a$$

The resultant transport velocity of each downstream vortex is V as shown in Fig. 17.

Solving for the velocity of propagation V_p from Fig. 17, one finds

$$V_p = C_{y2} - \frac{V_i}{2}\sin\theta - \left(C_{x1} - \frac{V_i}{2}\cos\theta\right)\tan\theta \left.\begin{array}{c} \\ \\ \end{array}\right\} \dots [23]$$

$$V_p = C_{y1} - C_{x1}\tan\theta$$

and if $\theta \approx 0$, $V_p \approx C_{y1}$.

This same result can be obtained easily by noting the parallelogram in the velocity diagram of Fig. 17 and solving for the equal and opposite side from V_p. Since the angle θ has been observed experimentally to be small in many cases, Equation [2] indicates that the fair experimental agreement between V_p and C_{y2} which has been noted by other observers is an immediate consequence. However, no assumptions regarding the angle θ are made in the analysis to follow.

To continue the analysis, the following variables are defined

$$N = \frac{V_i\cos\theta}{C_{z_2}}$$

$$M = \Gamma/\Gamma_0 = \frac{\text{shed vorticity}}{\text{bound vorticity}}$$

$$X = \frac{C_{z_1}}{C_{z_2}}$$

Y = fraction of downstream periphery covered by stall cell

$$= \frac{qa/\cos\theta}{d}$$

The circulation around a blade away from the cell is given by

$$\Gamma_0 = s(C_{y1} - C_{y2})\dots\dots\dots\dots[24]$$

Then

$$V_i = \frac{M\Gamma_0}{a} = M\frac{s}{a}(C_{y1} - C_{y2}) = \frac{NC_{z_2}}{\cos\theta}$$

$$\frac{s}{a} = \frac{N}{M(X\tan\beta_1 - \tan\beta_2)\cos\theta}\dots\dots\dots[25]$$

The time average continuity equation gives

$$dC_{z_1} = \left(d = \frac{qa}{\cos\theta}\right)C_{z_2} + \frac{qa}{\cos\theta}(C_{z_1} - V_i\cos\theta) \left.\begin{array}{c} \\ \\ \\ \end{array}\right\}$$

$$X - 1 = Y(1 - N) - Y \qquad\qquad \dots[26]$$

$$X = 1 - YN$$

Solving from Fig. 17 for the velocity with which the vortexes move along lines (b) and (c), one finds

$$V_a = \frac{C_{z_2}}{\cos\theta} - \frac{V_i}{2}$$

Since the frequency with which vortexes pass downstream points on lines (b) and (c) must be the same as the frequency with which vortexes are shed onto the lines

$$\frac{s}{a} = \frac{V_p}{V_a} = \frac{C_{y2} - C_{z_2}\tan\theta}{\dfrac{C_{z_2}}{\cos\theta} - \dfrac{V_i}{2}} = \frac{(\tan\beta_2 - \tan\theta)\cos\theta}{1 - \dfrac{N}{2}}\dots[27]$$

Eliminating s/a from Equations [25] and [27] gives

$$(\tan\beta_2 - \tan\theta)\cos^2\theta = \frac{N - N^2/2}{M(X\tan\beta_1 - \tan\beta_2)}\dots[28]$$

Now, in order to find θ, M and N must be known. For a completely analytical solution of the problem M and N would have to be obtained from the dynamic stall characteristics of the cascade airfoils. Fortunately, this difficulty can be circumvented because a wide range of experimental data from the cascade and rotor shows that both M and N are consistently close to unity

since the blades lose most of their lift and the axial velocity in the stall cell is small. In addition, for $N \approx 1$, solutions for θ from Equation [28] are very little affected by variations in N from 0.8 to 1.2. Consequently, θ is not very sensitive to the axial velocity in the stall cell. If $M = N = 1$, Equation [28] becomes

$$\tan \beta_2 - \tan \theta = \frac{1 + \tan^2 \theta}{2(X \tan \beta_1 - \tan \beta_2)}$$

which, when solved for θ and substituted in Equation [23] yields

$$\frac{V_p}{C_{x_2}} = X \tan \beta_1 \pm (X^2 \tan^2 \beta_1 - \tan^2 \beta_2 - 1)^{1/2} \ldots [29]$$

or in terms of C_{y_1} and C_{y_2}

$$\frac{V_p}{C_{y_1}} = 1 \pm \left[1 - \left(\frac{C_2}{C_{y_1}} \right)^2 \right]^{1/2} \ldots \ldots \ldots [30]$$

Of the two solutions, one gives a propagation velocity greater than C_{y_1} and implies that θ is a large negative angle; the other gives a propagation velocity smaller than C_{y_1} and a small value for θ. The first solution is discarded because the numerical analysis of the trajectory of the shed vortex mentioned earlier and all the experimental data show that V_p should be, and is, smaller than C_{y_1}. Consequently, the final result is

$$\frac{V_p}{C_{y_1}} = 1 - \left[1 - \left(\frac{C_2}{C_{y_1}} \right)^2 \right]^{1/2} \ldots \ldots \ldots [31]$$

If the flow outside the stall cell is assumed to be loss-free, then the static pressure rise across the stall cell is given by $\Delta p / \frac{1}{2} \rho C_1^2$ $= C_v = 1 - \left(\frac{C_2}{C_1} \right)^2$ and the propagation velocity may be written in the form

$$\frac{V_p}{C_{x_1}} = \tan \beta_1 - \left(\frac{C_p}{\cos^2 \beta_1} - 1 \right)^{1/2} \ldots \ldots \ldots [32]$$

Real values of propagation velocity are possible only if $C_p > \cos^2 \beta_1$. It is interesting to note that for the boundary case of $C_p = \cos^2 \beta_1$ the value of propagation velocity is precisely the same as that given by the small perturbation theory of section 2.3 for large wave lengths.

Careful hot-wire measurements of velocity have been made behind an isolated rotor to compare the predicted propagation velocity of Equation [31] with experimental values. Because of the absence of guide vanes, C_{y_1} is equal to the blade speed. In Fig. 18 the test values of V_p/C_{y_1} are compared with values predicted from Equation [31] using measured values of C_2 and with
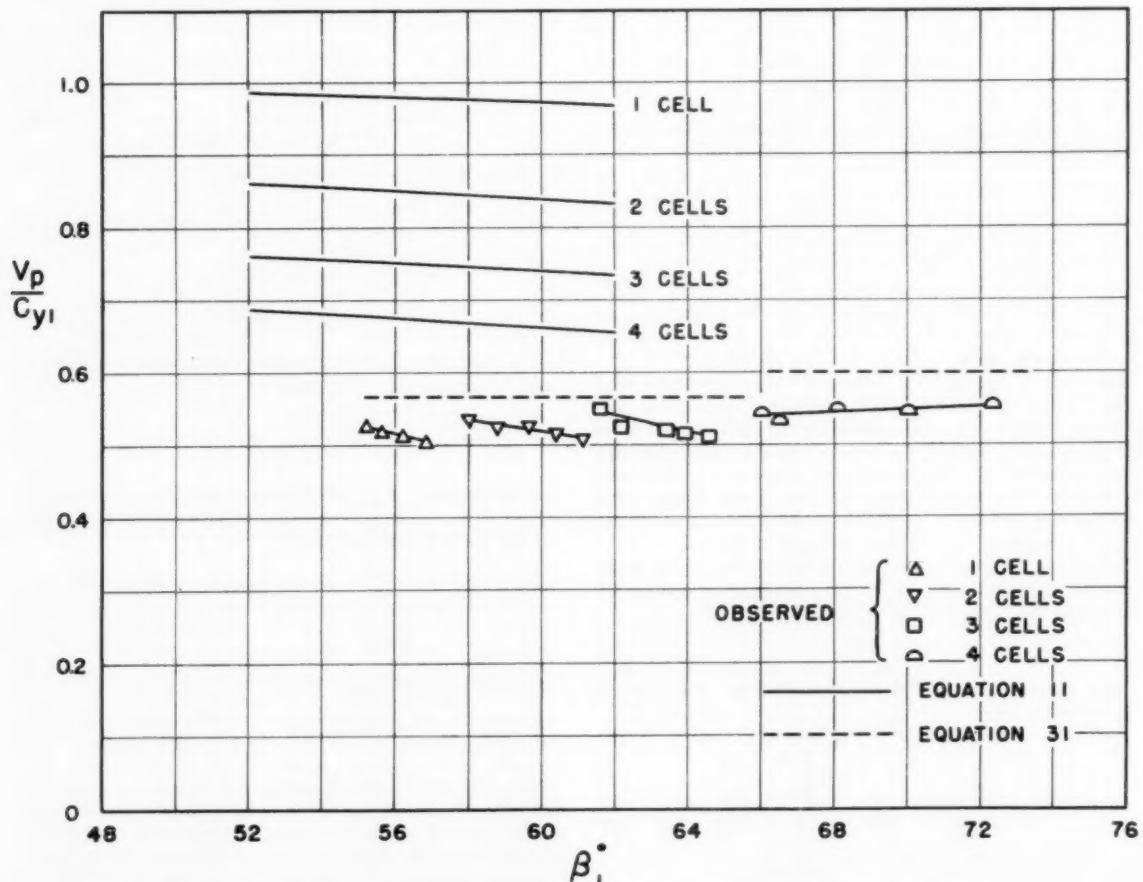

FIG. 18   THEORETICAL AND EXPERIMENTAL PROPAGATION VELOCITIES FOR AN ISOLATED ROTOR

the values predicted by the linearized theory. In this case, the small perturbation analysis of section 2.3 predicts propagation velocities which are considerably larger than the experimental values. The close agreement with Equation [31] suggests that the vortex model is worth studying in more detail close to the cascade.

An explanation for the variation of the number of stall cells in a cascade is also suggested by the vortex model. If the vortex trail does not extend infinitely far downstream but terminates close to the cascade, then the induced backflow at the center of the cell decreases as the width of the cell increases, until the airfoils in the center of the stall cell become unstalled and the stall cell splits. This effect may have some bearing on the small wave length of the stall cells in the circular cascade (for which the vortex trails extended only a few chord lengths downstream) and the larger wave length of the cells in the isolated rotor (for which the trails extended more than 20 chord lengths downstream).

BIBLIOGRAPHY

1 "Compressor Surge and Stall Propagation," by H W. Emmons, C. E. Pearson, and H. P. Grant, Trans. ASME, vol. 77, 1955, pp. 455–467.

2 "Experimental Investigations of Propagating Stall in Axial-Flow Compressors," by T Iura and W. D. Rannie, Trans. ASME, vol. 76, 1954, pp. 463–471.

3 "Some Stall and Surge Phenomena in Axial-Flow Compressors," by M. C. Huppert and W. A. Benser, *Journal of the Aeronautical Sciences*, vol. 20, December, 1953, pp. 835–845.

4 "Stall Propagation in Axial Compressors," by A. H. Stenning. A. R. Kriebel, and S. R. Montgomery, NACA TN 3580, June, 1956.

5 "Rotating Stall in Axial Compressors" (in English), by W. R. Sears, *Zeitschrift für angewandte Mathematik und Physik*, vol. 35, no. 6, 1955, pp. 429–455.

6 "Propagation of Stall in a Compressor Blade Row," by F. E. Marble, *Journal of the Aeronautical Sciences*, vol. 22, August, 1955, pp. 541–554.

7 "Fluid Mechanics," by V. L. Streeter, McGraw-Hill Book Company, Inc., New York, N. Y., 1951.

8 "Stall Propagation in a Cascade of Airfoils," by A. R. Kriebel, Gas Turbine Laboratory Report No. 36, Massachusetts Institute of Technology, Cambridge, Mass., August, 1956, to be published as NACA Technical Note.

9 "Mechanics of Deformable Bodies," by Arnold Summerfeld, Academic Press, New York, N. Y., 1950, pp. 231–238.

10 "Modern Developments in Fluid Mechanics," by S. Goldstein, Oxford University Press, London, England, vol. 1, 1952.

Appendix

CALCULATION OF α FOR A REAL CASCADE

Term α has been defined as the ratio of the actual mass flow through the cascade to the ideal mass flow for the same $(p_{01} - p_2)$ and flow deflection with no losses. It may be obtained from conventional test results as follows

$$\alpha = \frac{\text{actual mass flow}}{\text{ideal mass flow for same }(p_{01} - p_2)}$$

therefore

$$\alpha = \frac{c_1 \cos \beta_1}{(c_2)_{\text{no loss}} \cos \beta_2}$$

$$= \frac{c_1 \cos \beta_1}{\cos \beta_2 \left[\frac{2}{\rho}(p_{01} - p_2) \right]^{1/2}}$$

Since $p_{02} = p_{01}$ with no losses

$$\frac{c_1 \cos \beta_1}{\cos \beta_2 \left[\frac{2}{\rho}(p_{01} - p_1) - (p_2 - p_1) \right]^{1/2}}$$

Thus

$$\alpha = \frac{c_1 \cos \beta_1}{\cos \beta_2 (c_1^2 - C_p c_1^2)^{1/2}}$$

$$= \frac{\cos \beta_1}{\cos \beta_2 (1 - C_p)^{1/2}}$$

where

$$C_p = \frac{(p_2 - p_1)}{\frac{\rho}{2} c_1^2}$$

Discussion

ARTUR MAGER.[4] The writer commends the authors on the great amount of experimental and theoretical work done in connection with this important problem. Particularly, their experimental work carried out in stationary cascades and rotating blade row has resulted in valuable data and excellent photographs, which give a much better insight into the nature of the rotating stall than hitherto has been available.

With regard to the theoretical work however, the writer would like to raise a few objections and perhaps clarify some points. Let us concentrate first on the linearized solution. The authors have treated their problem as unsteady because they deal with a finite blade size and consequently they cannot consider the flow in a quasi-steady manner as is commonly done when the cascade is replaced by an actuating disk. However, the unsteady treatment of the problem carries with it some difficulties, which the authors seem to be bypassing. These difficulties concern the appropriate boundary conditions which the unsteady perturbation potential should satisfy at the cascade. While such boundary conditions are straightforward in the unstalled regions, e.g., the flow normal to the blades must vanish, what they should be in a stalled region is not clear. Now, the authors have used the conditions arising from the blocking effect of the stall region and from the inertia effect within the cascade, but these do not differentiate between the individual blades and, therefore Equation [8], for instance, does not involve the blade spacing, only the blade chord. This means that the applied boundary conditions do not insure that, in the unstalled region, the flow is necessarily adherent to the blades. To be consistent with these neglections one must imagine the blades as infinitesimally small, very closely spaced, or in other words we must again return to the actuator-disk concept. As a result of this objection, L/b should *always* be vanishingly small and consequently the troublesome term involving the harmonic number n in the expressions for the stall-propagating speed should be omitted. Equations [11] and [12] should therefore read

$$\frac{V_p}{c_s} = 2 \csc 2\beta_1(1 - C_p) \dots \dots \dots \dots [11]$$

$$\frac{V_p}{c_s} = \csc 2\beta_1(1 - C_p) \dots \dots \dots \dots [12]$$

[4] Research Scientist, Marquardt Aircraft Company, Van Nuys, Calif.

We note in passing that the authors have expressed considerable doubt as to the validity of Equation [12], basing it on the observed magnitude of the downstream pressure fluctuations.

In addition to the linearized solution, the authors also present an expression for the stall-propagation speed obtained by "nonlinear" considerations (Equation [31]). It may be interesting to dwell on the relation between Equations [11], [12], and [31]. Now, Equation [31] may also be rewritten after some manipulation as

$$\frac{V_p}{c_x} = \tan \beta_1 - \tan \beta_1 \left[1 - \frac{1 + \tan^2 \beta_1}{\tan^2 \beta_1} (1 - C_p) \right]^{1/2}$$

and when $C_p \to 1.0$, that is when the losses are very small, but the work input is large, one can linearize the square root to get Equation [12]

$$\frac{V_p}{c_x} \Bigg]_{C_p \to 1.0} = \frac{1 + \tan^2 \beta_1}{2 \tan \beta_1} (1 - C_p) = \csc 2\beta_1 (1 - C_p)$$

It is significant to note then, in view of the fact that the authors have stressed the validity of Equation [11] in preference to Equation [12], that this is not consistent with their own result given by the nonlinear theory.

The limiting value of C_p in Equation [31] of the paper is $C_p = \cos^2 \beta_1$ and for this limiting value, Equation [11] agrees exactly with Equation [31]. It is therefore appropriate to ask what does this special variation of C_p imply. Solving this statement for p_2, one obtains

$$p_2 = p_1 + \frac{1}{2} \rho C_x^2 = P_1$$

Or in other words, in absence of prerotation, a condition like this occurs when the downstream static pressure equals the absolute upstream stagnation pressure. Now remembering that in the nonlinear model of stall the flow is completely blocked by the stalled regions, and there is no work done on the fluid in these regions, one realizes that a situation like this would have to exist when the rotor is completely stalled. Incidentally, for a stationary cascade, with prerotation, $C_p = \cos^2 \beta_1$, implies that at the cascade all of the energy of the inlet flow has been converted into that connected with a purely whirling flow, which, in the nonlinear model would again happen when the cascade is completely stalled. The occurrence of the limiting value of C_p is therefore quite consistent with the postulated stall model and the agreement with Equation [11] which was derived by neglection of stall effects on the downstream field is quite understandable.

Authors' Closure

The authors wish to thank Dr. Mager for his stimulating comments and discussion.

For the small perturbation analysis, we agree that the blades must be very closely spaced so that α is a continuous function of y. However, it is not necessary that L should be vanishingly small and Equations [11] and [12] are in fact valid as written in their original form. We have found no way to rationalize the troublesome harmonics out of the analysis. It is possible that the propagation velocity should be regarded as the "group" speed of a number of wave trains, rather than the speed of a single component.

Comparison of the linearized and large disturbance theories is unprofitable, since the models are completely different. There is no reason for the two theories to give the same result, and we did not intend that any conclusions should be drawn from the coincidence of propagation speeds for one special condition.

Equation [31] can be derived in a completely different way,[5] on the assumption that the stalled zone is equivalent to a flat plate moving along the cascade in equilibrium under the forces acting on it. When approached in this way, the significance of the critical value of C_p becomes apparent, as this is the smallest value of C_p for which the force balance is possible.

[5] "Stall Propagation in a Cascade of Airfoils," by A. H. Stenning, B. S. Seidel, and Y. Senoo, Gas Turbine Laboratory Report No. 41, August, 1957, under NACA contract NAw6484.

Losses in Flow Normal to Plane Screens

By W. G. CORNELL,[1] EVENDALE, OHIO

Experimental data are correlated for total-pressure losses in subsonic-compressible, real-fluid flow normal to plane round-wire screens of square mesh, including new data at very high screen Reynolds number. A theory is developed for subsonic-compressible flow normal to plane screens of sharp-edged elements of any geometry, yielding predicted total-pressure losses which compare reasonably well with experimental data. The effect of Reynolds number on loss is found to be small, except at very low Reynolds number.

Nomenclature

THE following nomenclature is used in the paper:

a = velocity of sound = $(\gamma gRT)^{1/2}$ for a perfect gas

A_b = "blocked area" or projected area of elements on plane parallel to screen, sq ft

A_o = "open area" or projected area of openings on plane parallel to screen, sq ft

d = diameter of screen wire, ft

g = acceleration due to gravity, ft/sec^2

M = Mach number = w/a

p = static pressure, psf

p_T = total or impact pressure, psf

R = gas constant for perfect gas, ft/deg R

Re = Reynolds number

s = $A_b/(A_b + A_o)$, solidity or ratio of blocked area to total area

T = absolute temperature, deg R

w = velocity of fluid, fps

β = $\rho w/\rho_T a_T$

γ = specific heat ratio of perfect gas

δ = diameter of holes in perforated plate, ft

Δp_T = total-pressure loss in screen, psf

ϵ = $M[1 + (\gamma - 1)M^2/2]^{1/2}/(1 + \gamma M^2)$

λ = $\Delta p_T/(1/2)\rho_1 w_1^2$

ν = kinematic viscosity of fluid, ft^2/sec

ξ = strip width of strip screens, ft

ρ = mass density of fluid, slug/cu ft

φ = contraction coefficient, or ratio of jet area to orifice area

Subscripts

1—refers to flow condition ahead of screen

2—refers to flow condition in jet behind screen

3—refers to flow condition after mixing behind screen

T—refers to total or impact conditions

Introduction

The flow of fluids in screens is of practical interest in many engineering problems, e.g., removal of foreign objects from jet-engine inlet flow, simulation of nonuniform inlet-flow distributions for jet engines, smoothing flow, and producing turbulence in wind tunnels, and so on. In such cases, one problem of interest

is the prediction of the total-pressure loss caused by the screen in terms of screen geometry and upstream-flow conditions.

The losses in screen flow are evidently determined by screen geometry, scale effect (Reynolds number), and compressibility effect (Mach number). The literature contains considerable information on screen-geometry effect, some on Reynolds-number effect, and little on Mach-number effect. It is the purpose of the present investigation to assemble the available information, supplementing where necessary, and to present a consistent and inclusive method for prediction of losses in subsonic-compressible, real-fluid flow normal to plane screens composed of either round wires in square-mesh array or sharp-edged elements, such as thin strips or lands between openings in perforated sheets, in any array. Flow in other screen configurations is discussed.

Geometry of Screens

A screen may be defined as a regular assemblage of elements forming a pervious sheet which is relatively thin in the direction of flow through the screen. Examples are woven round-wire screens, perforated thin sheets, grids of bars of rectangular cross section, screens composed of streamlined wire, and so on. Porous media such as drilled thick plates, sintered aggregates, packed-particle beds, multiple layers of screens, and so on, are not included since they are relatively thick in the flow direction; viz., are characterized by a large ratio of length to breadth of the internal flow path.

A screen is characterized geometrically by element type (round-wire, etc.), by element arrangement (square mesh, etc.), and by screen shape (plane, etc.). One convenient nondimensional parameter is some measure of the area blocked by screen elements as a fraction of total available flow area. In the case of flow normal to plane screens, such a parameter of solidity s may be defined as the ratio of blocked area A_b to total area $A_b + A_o$, where A_b is the projected area of the screen elements and A_o that of the openings, projections being taken on a plane parallel to the plane of the screen.

In the case considered, the screen is taken to be across the entire flow; any "end effects" of finite extent of screen relative to flow extent being neglected.

Mechanism of Screen Flow

The flow over a screen is characterized by average upstream Mach number M_1, by a suitably defined screen element Reynolds number Re, and by inclination of the upstream flow to some characteristic dimension of the screen assemblage. The flow is accelerated through the screen and forms jets of relatively high velocity behind the openings, interspersed with wakes of relatively low velocity behind the elements. As the flow passes downstream, the jets and wakes mix to yield finally a uniform flow. Most of the total-pressure loss occurs in the mixing process, although some loss occurs in the flow within the screen passages, particularly in the case of round-wire and similar screens. This latter loss is small in sharp-edged screens.

Theoretical prediction of screen loss requires prediction of the loss within the screen passages and of the downstream mixing losses. The former could be approximated in the case of round-wire screens, but the latter loss depends upon the relative extent of the jets and wakes, a difficult quantity to predict. Hence, the round-wire screen loss is best determined empirically. The

sharp-edged screen loss can be attacked theoretically, since internal flow losses are small and relative extent of jets and wakes can be predicted since the sharp edges determine the points of jet formation.

A convenient measure of the total-pressure loss is the non-dimensional loss coefficient $\lambda = \Delta p_T/(1/2)\rho_1 w_1{}^2$, based on the upstream dynamic pressure. The problem may be stated as follows: Given a plane screen of certain element type and element arrangement, with a given upstream flow normal to the screen, it is desired to find the downstream-flow conditions, in particular, the total pressure. Functionally, it is desired to obtain the loss coefficient $\lambda_{(s,\ Re,\ M_1)}$.

Low Mach-Number Flow Normal to Plane Round-Wire Screens of Square Mesh

As previously discussed, round-wire screen losses are more amenable to empirical prediction than to theoretical analysis. An excellent correlation method is the method of Wieghardt (1),[2] who visualized the flow as approximately that over single infinite cylinders bathed in a uniform flow of velocity $w_1/(1 - s)$, due to the constriction imposed by the screen. Thus, Wieghardt cor-

[2] Numbers in parentheses refer to the Bibliography at the end of the paper.

related $\lambda(1 - s)^2/s$ with $Re = w_1 d/(1 - s)\nu_1$, anticipating a curve similar in trend to the drag coefficient—Reynolds number curve for single infinite cylinders, and hoping to minimize the effect of s as a parameter. The loss factor $\lambda(1 - s)^2/s$ can be shown (1) to be equal to the average screen-wire drag coefficient, based on the velocity $w_1/(1 - s)$. Wieghardt's correlation covers the range $60 < Re < 1000$ and is represented by the relation

$$\lambda(1 - s)^2/s = 6\ (Re)^{-1/3} \dotfill [1]$$

MacDougall (2) correlated data in the range $0.006 < Re < 20$ and obtained the relation

$$\lambda = \frac{33.93}{Re}\ \frac{s(1 - s)^{-1.77}}{1 + (1 - s)^{1/2}} \dotfill [2]$$

The present correlation uses Wieghardt's variables $\lambda(1 - s)^2/s$ and $w_1 d/(1 - s)\nu_1$ and correlates data of references (4-11) for flow of various liquids and air as well as the present data on air flow. The range covered is $40 < Re < 17{,}000$. The results of the correlation are shown in Fig. 1, compared to the dashed line of Wieghardt's correlation and to the single drag curve (3). The correlation is shown in Fig. 2 the shaded area indicating the range of experimental data. MacDougall's low Re results, as expressed by Equation [2], have been faired in to the single
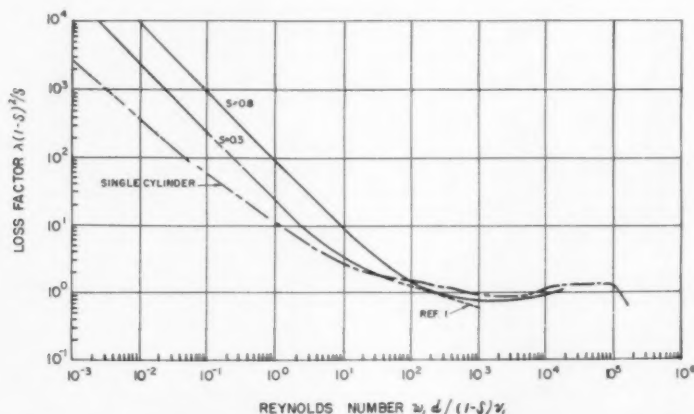
Fig. 1 Comparison of correlation of experimental round-wire-screen losses and of single-cylinder drag coefficients in low-velocity flow
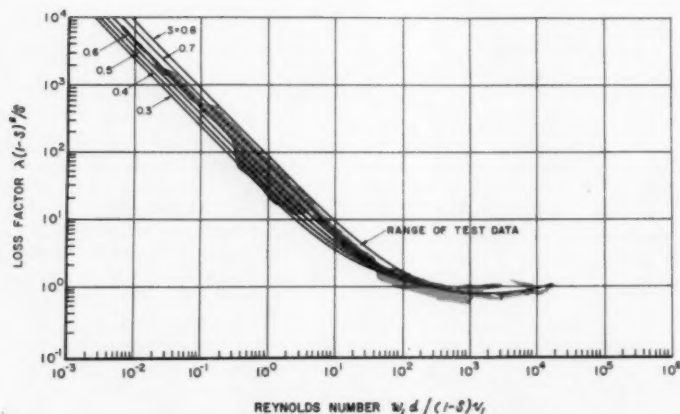
Fig. 2 Correlation of experimental round-wire-screen losses in low-velocity flow

curve taken to correlate the various high Re data. Fig. 3 shows data points in the high Re range, compared with the correlation. The points shown as triangles are the present data on screen No. 1, which had a rectangular mesh of side ratio 19/16 Consequently, the loss values may be relatively low compared to those of square-mesh screens, according to Wieghardt (1). It can be seen that the $\lambda(1 - s)^2/s$ curve does not fall continuously as Re is increased, but rather levels off and rises again as Re is increased to the range of Re = 17,000. This behavior would be expected on the basis of the single-cylinder model and has been discussed as a possibility by Annand (11). It may be hypothesized that for Re above 17,000 the curve ultimately should flatten, drop precipitously at some critical Reynolds number, and continue at a lower level. Further data at higher Re are needed in order to check this hypothesis. The screen and single-cylinder curves show the same general trends at the higher Reynolds numbers, exact correspondence not being expected owing to neglected effects of screen-mesh corners, interference between wires, and so on. At low Reynolds numbers, the screen-data

spread requires the introduction of s as a parameter, as found by MacDougall (2).

High Subsonic Mach-Number Flow Normal to Plane Round-Wire Screens of Square Mesh

The compressibility effect in high-speed gas flow through screens becomes large at high inlet Mach number M_1, especially at high solidity s, since very high velocities then occur within the screen passages. Experimental results of Adler (4) are shown in Fig. 4, based on high-velocity air-flow tests. The loss coefficient λ is seen to rise with M_1 for a given s, as might be expected physically. Adler does not report Reynolds-number data, but extrapolation of his data to $M_1 = 0$ (incompressible flow) gives values comparable to those of the flat, high Re range of the present correlation, albeit somewhat lower than the average curve. Thus, the magnitudes of Adler's values may be low, but the variations with M_1 and s are clearly shown, so that reasonable values of $\lambda(M_1)/\lambda(0)$ can be readily estimated. The value of $\lambda(0)$ for low-velocity flow can then be obtained from the correlation
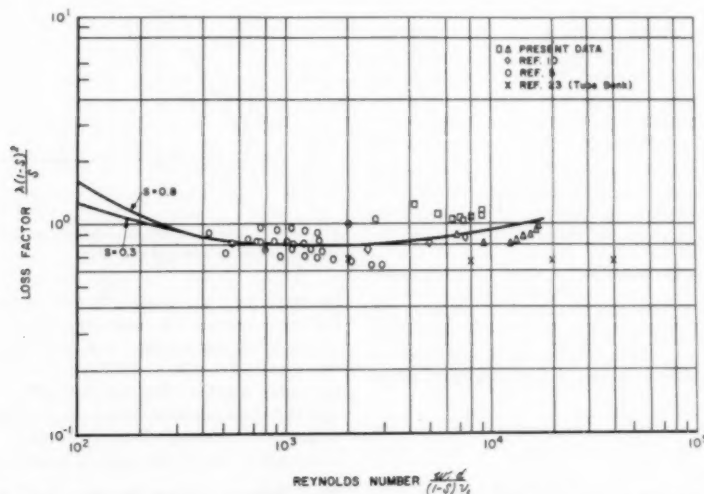
Fig. 3 Experimental round-wire-screen losses in low-velocity, high Reynolds number flow
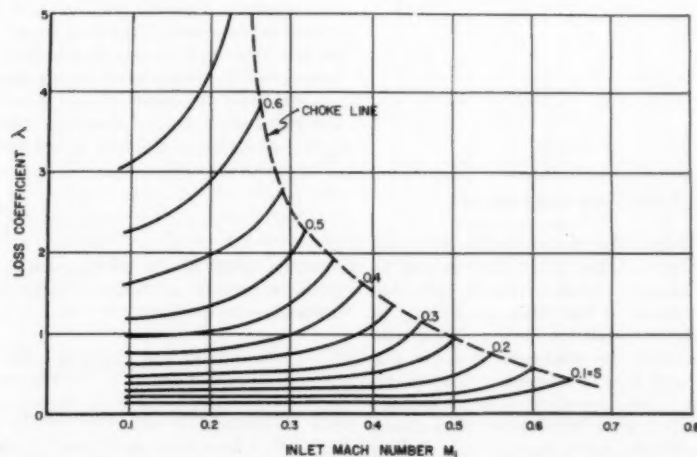
Fig. 4 Experimental round-wire-screen losses in high-velocity flow, after Adler (4)

of Fig. 2. The empirical data do not cover the combination of high-velocity low-Reynolds number flow. However, the suggested procedure should give at least first-order results in such cases, in view of the known effects of compressibility and friction on the drag of various bodies and the loss in various systems.

The choke limit shown in Fig. 4 is the locus of "choked flow" conditions, in the ordinary sense of choking a rounded flow nozzle, viz., for fixed upstream total pressure, decreasing downstream static pressure increases flow until choking occurs, beyond which point flow remains constant unless upstream total pressure changes. Adler (4) shows that the experimental choke limit agrees very well with a theoretical prediction based on the value of M_1 required to produce sonic velocity in the effective minimum area of the screen; viz., $(1 - s)$ times inlet area.

Low Mach-Number Flow Normal to Plane Sharp-Edged Screens

As previously discussed, a theoretical approach is possible in the case of sharp-edged screens. Fig. 5 shows a two-dimensional model of the flow around one half of a screen element, the model being defined as having the same solidity $s = A_b/(A_b + A_o)$ as the equivalent screen. The model may be thought of as a cascade of sharp-edged flat plates, in analogy to the round-wire screen model, or as a succession of sharp-edged orifices of area A_o placed in channels of area $A_b + A_o$. Jets of area φA_o are found downstream, interspersed by wakes of area $A_b + A_o(1 - \varphi)$. The relative extent of jets and wakes is characterized by solidity s and contraction coefficient $\varphi$. Weinig (12) used this model for the flow in ribbon parachutes and compared the results successfully with experimental data of Flachsbart (5) for losses in strip screens, showing that mixing of the jets and wakes to form a uniform stream gives rise to a loss coefficient given by

$$\lambda = [1 - \varphi(1 - s)]^2/\varphi^2(1 - s)^2 \dots \dots [3]$$

under the assumptions of incompressible, perfect-fluid flow, the wake velocity being taken as zero. Weinig used the theoretical results of von Mises (13) for contraction coefficient φ, based on a free-streamline potential-theory model of flow in a sharp-edged orifice.
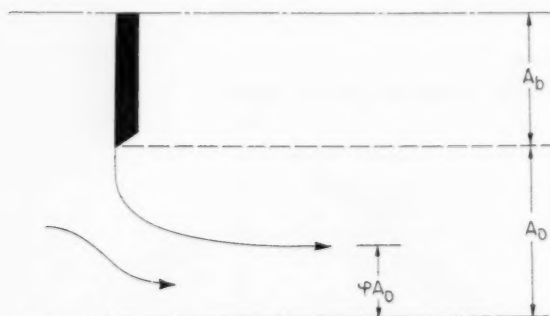
Fig. 5 Model for flow in sharp-edged screens

The method may be applied to other sharp-edged screens, such as perforated plates. Further, a better approximation may be obtained, as follows: It has been noted by Rouse (14), that von Mises' theory underestimates φ somewhat, particularly at low s. Using Rouse's empirical correlation for $\varphi(s)$, along with the relation [3] for λ, there results the solid curve shown in Fig. 6, compared with experimental data of references (5, 9, 15–19) as well as the present data. The dashed curve shows Weinig's theory. The experimental data show lower losses than theory would predict, undoubtedly due in part to imperfect sharpness of test screen edges. The effect of Reynolds number was found to
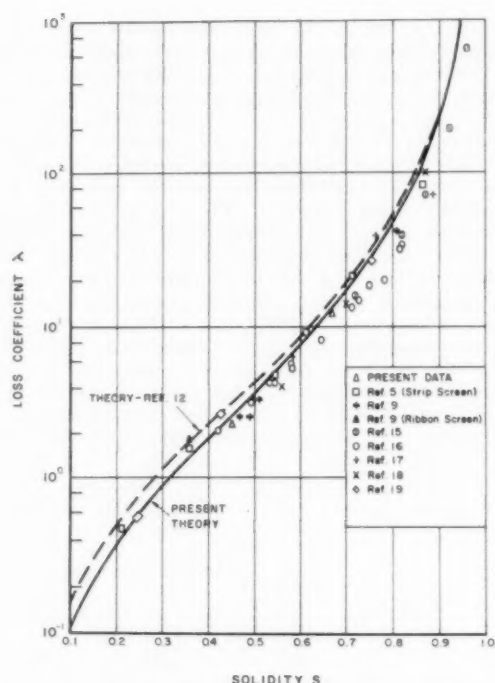
Fig. 6 Comparison of theoretical and experimental sharp-edged screen losses in low-velocity flow

be small over the range covered by the various investigators. For strip screens, the range was $250 < w_1\xi/(1 - s)\nu_1 < 21,000$, where ξ is strip width. For perforated plates, the range was $400 < w_1\delta/(1 - s)\nu_1 < 44,000$, where δ is hole diameter. At low Reynolds number, the loss coefficient may be expected to fall and then rise again as Reynolds number is decreased, based on orifice contraction-coefficient behavior. Some indication of such a trend is shown by the data of Taylor and Davies (9).

High Subsonic Mach-Number Flow Normal to Plane Sharp-Edged Screens

As in the case of round-wire screens, compressibility effect becomes large at high M_1 and high s of sharp-edged screens. A theoretical approach is possible, by generalizing the analysis of the low Mach-number case, if the effect of M_1 on φ can be estimated, and by compressible mixing calculations.

Considering the model of Fig. 5 in the case of compressible flow of a perfect gas and denoting, respectively, by indexes 1, 2, 3, the upstream, jet, and fully mixed conditions, the conservation of mass requires that

$$\frac{\rho_1 w_1}{1 - s} = \rho_2 w_2 \varphi = \frac{\rho_2 w_3}{1 - s} \dots \dots [4]$$

Assuming the flow in the wakes at station 2 to be at rest and at the jet pressure p_2, the conservation of momentum in the mixing process requires that

$$\frac{p_2}{1 - s} + \rho_2 w_2^2 \varphi = \frac{p_3}{1 - s} + \frac{\rho_2 w_3^2}{1 - s} \dots \dots [5]$$

The equation of state of a perfect gas yields

$$\frac{p_1}{\rho_1 T_1} = \frac{p_2}{\rho_2 T_2} = \frac{p_3}{\rho_2 T_3} \dots \dots [6]$$

Assuming zero loss in the jet formation, viz., $p_{T1} = p_{T2}$, noting that in absence of heat transfer total temperature $T_{T1} = T_{T2} = T_{T3}$, where $T_T/T = 1 + (\gamma - 1)M^2/2$, introducing the Mach number, $M = w/a = w/(\gamma gRT)^{1/2}$, the mass flow function β, defined by

$$\beta = \frac{\rho w}{\rho_T a_T} = M[1 + (\gamma - 1)M^2/2]^{\frac{\gamma+1}{2(1-\gamma)}} \dots \dots [7]$$

and a compressible-flow function defined by

$$\epsilon = M[1 + (\gamma - 1)M^2/2]^{1/2}/(1 + \gamma M^2) \dots \dots [8]$$

the Relations [4], [5], and [6] may be combined to yield

$$\frac{\beta_1}{\beta_2} = (1 - s)\varphi \dots \dots \dots \dots [9]$$

$$\frac{\epsilon_2}{\epsilon_3} = \frac{1 + \gamma M_2^2\varphi(1 - s)}{(1 + \gamma M_2^2)\varphi(1 - s)} \dots \dots \dots [10]$$

$$\frac{p_{T3}}{p_{T1}} = \frac{\beta_2}{\beta_3}(1 - s)\varphi \dots \dots \dots [11]$$

The Relations [9], [10], and [11] may then be solved for the total-pressure ratio p_{T3}/p_{T1} as follows: Given a screen of solidity s and inlet Mach number M_1, the contraction coefficient φ is determined as discussed later. Use of M_1 in Equation [7] yields β_1, which with s and φ in Equation [9] yields β_2. Then, iterative or graphical use of β_2 in Equation [7] yields M_2, which is substituted in Equation [8] to yield ϵ_2. Then, use of ϵ_2, M_2, φ, and s in Equation [10] yields ϵ_3. Then, iterative or graphical use of ϵ_3 in Equation [8] yields M_3. Finally, use of M_3 in Equation [7] yields β_3, which with β_2, s, and φ may be substituted in Equation [11] to yield the total-pressure ratio p_{T3}/p_{T1}.

The total-pressure loss coefficient may then be calculated as

$$\lambda = \frac{p_{T1} - p_{T3}}{1/2\rho_1 w_1^2} = \frac{1 - p_{T3}/p_{T1}}{(\gamma M_1^2/2)[1 + (\gamma - 1)M_1^2/2]^{-\gamma/(\gamma-1)}} \dots [12]$$

utilizing the ordinary adiabatic relationships for a perfect gas.

In order to obtain the contraction coefficient φ for chosen s and M_1, recourse must be had to experimental data, since no general theoretical treatment is at hand. Thus, the test results of Grey and Wilsted (20) were used. Experimental discharge coefficients for air flow in sharp-edged orifices, along with an approximate experimental velocity coefficient of 0.928, were used to obtain the variation of φ with M_1 and s. The smoothed test results of Grey and Wilsted were presented as curves of discharge coefficient versus total to static-pressure ratio across the orifice. These results were adjusted to give $\varphi = 1$ at $s = 0$ by dividing all discharge coefficients by 0.928, a value not far from the experimentally determined velocity coefficient, which was not strongly affected by pressure ratio or orifice-area ratio. Finally, there resulted the curves of Fig. 7, where φ is plotted versus p_{T1}/p_2 for various s. In applying the results, the abscissa is used as $p_{T2}/p_2 = [1 + (\gamma - 1)M_2^2/2]^{\gamma/(\gamma-1)}$, since no loss is assumed in the process of jet formation. Thus, Fig. 7 gives $\varphi(M_1, s)$ implicitly, taking into account the remainder of the calculation of λ.

The loss coefficient is then obtained as a function of s and M_1 from Equation [12] and plotted as shown in Fig. 8. A comparison of the values of λ at $M_1 = 0$ with the predicted incompressible curve of Fig. 6 shows that exact correspondence is not obtained, since different empirical data were used to formulate the two predictions. However, the disparity is not large and the

curves of Fig. 8 can be used effectively to predict $\lambda(M_1)/\lambda(0)$. The value of $\lambda(0)$ is obtained from Fig. 6.

The "subsonic flow limit" curve is the locus of conditions at which the jet velocity w_2 is sonic. The "upper bound for choke" curve corresponds to sonic velocity in the screen openings. Actual choking conditions will obtain somewhere between the two limit curves, but no general method for estimation of these

Fig. 7 Semi-empirical sharp-edged pipe-orifice contraction coefficients in high-velocity flow

Fig. 8 Theoretical sharp-edged screen losses in high-velocity flow
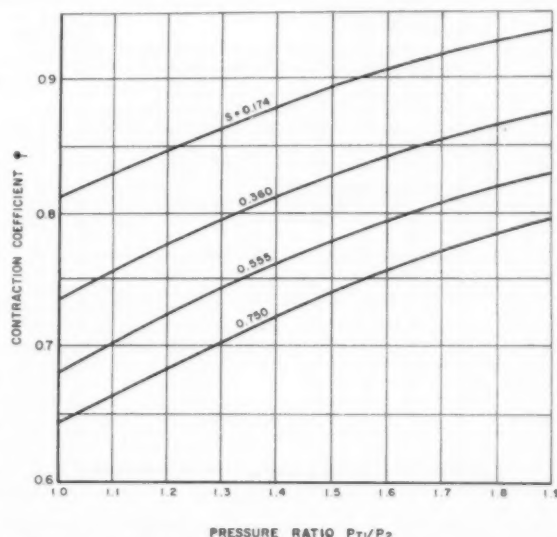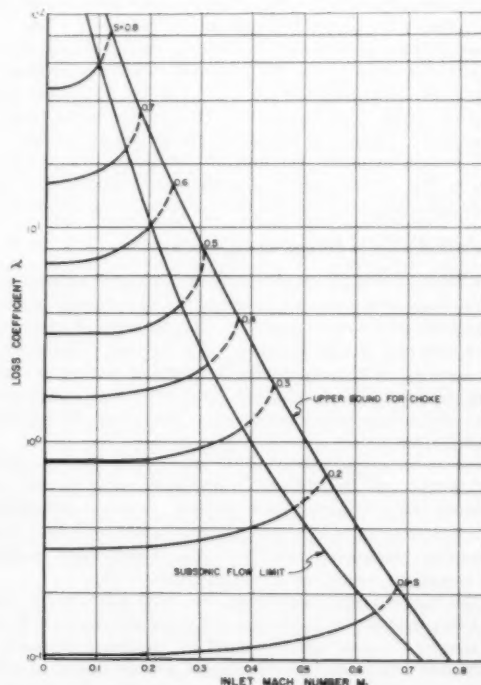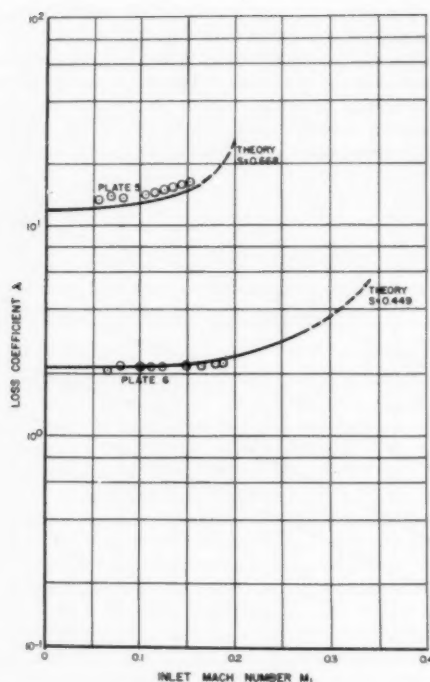
Fig. 9 Comparison of theoretical and experimental sharp-edged screen losses in high-velocity flow

conditions is at hand. The loss-coefficient curves are dashed between the two limit curves, since the one-dimensional basis of the present prediction method cannot be used for $M_2 > 1$. The mechanics of the method given herein give no physically allowable solution for $M_2 > 1$; viz., the predicted total-pressure drop decreases as M_2 is increased above unity. Further, additional supersonic-flow losses, notably resulting from shock waves, are not accounted for. Thus, the dashed curves are extrapolations.

The present experimental data on two perforated plates are shown in Fig. 9, compared to the theoretical prediction. Reasonable agreement is found between experiment and theory.

Further Problems of Screen Aerodynamics

The effect of element arrangement or mesh type appears to be small in sharp-edged screens, based on the wide variety of model screens used in the tests analyzed. In the case of round-wire screens, this may not be the case. For instance, Wieghardt (1) notes a set of tests on a round-wire screen of rectangular mesh having side ratio 1.25 gave rather lower losses than square-mesh screens at the same Reynolds number. The present data support this point of view.

Losses in screens composed of elements other than those discussed herein will be different. Streamline-wire screens will have lower losses than round-wire screens. Losses in rectangular-bar screens of more or less square-edged variety will have losses intermediate between round-wire and sharp-edged screens. Data on various screens are given by Hoerner (21).

In the case of staggered screens, viz., with inlet flow at some angle other than 90 deg to the plane of the screen, losses will vary with stagger. Losses are decreased by moderate amounts of stagger, say, up to 60 deg. Staggered screens deflect the flow in the same fashion as a diffusing cascade, the flow being refracted towards the normal to the plane of the screen. Data

on staggered screens are given by Simmons and Cowdrey (10) and Hoerner (21).

When a screen extends only partially across a channel, end effects appear to decrease the loss, as might be expected. Flachsbart (5) presents data on screens tested both in pipes and free airstreams.

In the case of multiple or packed screens, the mechanism of the flow becomes more like that in porous media such as sintered aggregates. Loss data for flow in packed screens are given by Grootenhuis (7) and Tong (22).

Present Experimental Data

The test apparatus consisted of a turbosupercharger-type

Table 1 Data for Screen and Plates Tested

Round-Wire Screens

Screen no.	Wire diam, in.	Wire center-line spacing, in.	Solidity
1	0.073	0.160 × 0.190	0.665
2	0.089	0.515 × 0.515	0.314

Perforated Plates

Plate no.	Hole diam, in.	Hole center spacing between rows, in.	Solidity
5	3/16	0.312 × 0.266	0.668
6	1/4	0.312 × 0.266	0.449[a]

[a] 3.2 per cent of open area made up of 3/16-in.-diam. holes near pipe wall.

Table 2 Test Results

Screen no.	Solidity	$\dfrac{w_1 d}{(1-s)p_1}$	M_1	λ	$\dfrac{\lambda(1-s)^2}{s}$
1	0.665	6900	0.0579	5.34	0.901
		9300	0.0797	4.83	0.814
		11400	0.1009	4.82	0.813
		12500	0.1154	4.81	0.812
		13300	0.1281	4.91	0.830
		14200	0.1425	5.09	0.859
		15900	0.1591	5.26	0.886
		16500	0.1696	5.59	0.943
		17000	0.1834	5.83	0.984
2	0.314	4200	0.0623	0.834	1.251
		5500	0.0859	0.749	1.123
		6500	0.1059	0.710	1.068
		7100	0.1204	0.723	1.087
		7400	0.1355	0.719	1.080
		8100	0.1514	0.724	1.088
		8500	0.1750	0.723	1.087
		8900	0.1880	0.736	1.103
		8900	0.2029	0.770	1.158

Plate no.	Solidity	$\dfrac{w_1 \delta}{(1-s)p_1}$	M_1	λ
5	0.668	18000	0.0561	13.64
		22100	0.0697	14.90
		28300	0.0897	14.21
		33300	0.1047	14.29
		37300	0.1183	14.92
		39200	0.1267	15.50
		42100	0.1371	15.90
		43500	0.1442	16.59
		44100	0.1507	17.67
6	0.449	16700	0.0663	2.18
		20100	0.0815	2.36
		24700	0.1034	2.25
		27000	0.1169	2.26
		28400	0.1317	2.25
		32200	0.1514	2.26
		33100	0.1679	2.29
		35200	0.1814	2.34
		35400	0.1897	2.39

centrifugal compressor, driven by an electric motor, discharging into a plenum chamber containing a multiple nozzle plate for flow measurement. The plenum chamber discharged into a 9-in. pipe, in which the test screens were mounted between flanges. Screen pressure drop was measured by pipe-wall static taps in planes 6 in. up and downstream of the screen plane. An additional length of pipe discharged to atmosphere.

The screens tested are described in Table 1. The experimental results are given in Table 2.

Conclusion

The present investigation yields methods for predicting total-pressure losses for subsonic-compressible, real-fluid flow normal to plane screens.

For the case of low Mach-number flow normal to plane round-wire screens of square mesh, Fig. 2 gives loss coefficient λ as a function of solidity s and Reynolds number Re. For high subsonic Mach-number flow, Fig. 4 gives $\lambda(M_1)/\lambda(0)$ as a function of s and M_1, to be used in connection with $\lambda(0)$ from Fig. 2.

Similarly, for the case of low Mach-number flow normal to plane sharp-edged screens, Fig. 6 gives λ as a function of s. For high subsonic Mach-number flow, Fig. 8 gives $\lambda(M_1)/\lambda(0)$ as a function of s and M_1, while Fig. 6 gives $\lambda(0)$. Reynolds number Re has relatively little effect on loss coefficient λ, except for very low Re.

Acknowledgments

The author is indebted to Arthur Sherman and Kenneth Hanson for helpful discussions, suggestions, and aid in the present investigation and to Mrs. Eunice Hart for assistance with calculations and preparation of the manuscript. Many of the data given herein have previously been submitted in concise form for publication in the second edition of the Reactor Handbook of the Atomic Energy Commission.

Bibliography

1 "On the Resistance of Screens," by K. E. G. Wieghardt, *The Aeronautical Quarterly*, vol. 4, February, 1953, pp. 186–192.
2 "Pressure Drop Through Screens," by D. A. MacDougall, MS thesis, Ohio State University, Columbus, Ohio, 1953.
3 "Modern Developments in Fluid Dynamics," by S. Goldstein, Oxford University Press, London, England, 1938, p. 419.
4 "Variations with Mach Number of Static and Total Pressures Through Various Screens," by A. A. Adler, NACA Wartime Report L-23 (originally Confidential Bulletin L5F-28, 1946).
5 "Widerstand von Seidengazefiltern, Rundraht- und Blechsstreifensieben mit quadradischen Maschen," by O. Flachsbart, Ergebnisse der Aerodynamischen Versuchsanstalt zu Goettingen, Lieferung IV, 1932, pp. 112–118.
6 "The Aerodynamic Characteristics of Damping Screens," by G. B. Schubauer, W. G. Spangenberg, and P. S. Klebanoff, NACA TN 2001, January, 1950.
7 "A Correlation of the Resistance to Airflow of Wire Gauzes," by P. Grootenhuis, Proceedings of The Institution of Mechanical Engineers, vol. 168, no. 34, 1954, pp. 837–846.
8 "The Effect of a Gauze on the Velocity Distribution in a Uniform Duct," by A. R. Collar, Aeronautical Research Council, Reports and Memoranda 1867, 1939.
9 "The Aerodynamics of Porous Sheets," by G. I. Taylor and R. M. Davies, Aeronautical Research Council, Reports and Memoranda 2237, 1944.
10 "Measurements of the Aerodynamic Forces Acting on Porous Sheets," by L. F. G. Simmons and C. F. Cowdrey, Aeronautical Research Council, Reports and Memoranda 2276, 1949.
11 "The Resistance to Air Flow of Wire Gauzes," by W. J. D. Annand, *Journal of the Royal Aeronautical Society*, vol. 57, March, 1953, pp. 141–146.
12 "Parachutes with Canopies Composed of Self-Supporting Ribbons," by F. S. Weinig, Hq. AMC, Wright Field, Technical Report F-TR-2148-ND, GS-AAF-Wright Field 22, 1947, pp. 13–14.
13 "Berechnung von Ausfluss- und Überfallzahlen," by R. von Mises, *Zeitschrift des Vereines deutscher Ingenieure*, vol. 61, no. 21, 1917, pp. 447–452; no. 22, pp 469–474; no. 23, pp. 493–498.
14 "Elementary Mechanics of Fluids," by H. Rouse, John Wiley & Sons, Inc., New York, N. Y., 1946, pp. 58–59.
15 "Wind-Tunnel Experiments on the Squidding of Parachutes," by L. F. G. Simmons, R. W. F. Gould, and C. F. Cowdrey, Aeronautical Research Council, unpublished Report ARC 7062, November, 1943.
16 "Pressure-Drop Characteristics of Orifice Plates Used to Simulate Radiators," by K. R. Czarnecki, NACA Wartime Report L-342 (originally Advance Restricted Report, March, 1942).
17 "The Resistance to Air Flow of Porous Materials Suitable for Boundary-Layer Control Applications Using Area Suction," by R. E. Dannenberg, J. A. Weiberg, and B. J. Gambucci, NACA TN 3094, 1954.
18 "Full-Scale Wind-Tunnel Investigation of Wing Cooling Ducts," by F. R. Nickle and A. B. Freeman, NACA, ACR, October, 1938.
19 "An Investigation of Flow Through Screens," by W. D. Baines and E. G. Peterson, Trans. ASME, vol. 73, 1951, pp. 467–480.
20 "Performance of Conical Jet Nozzles in Terms of Flow and Velocity Coefficients," by R. E. Grey and H. D. Wilsted, NACA TN 1757, November, 1948.
21 "Pressure Losses Across Screens and Grids," by S. F. Hoerner, USAF Technical Report 6289, Wright Air Development Center, November, 1950.
22 "Heat Transfer and Friction Characteristics of Screen Matrices at High Reynolds Number," by L. S. Tong, Stanford University, Department of Mechanical Engineering, Stanford, Calif., Technical Report 28, April, 1956.
23 "Correlation and Utilization of New Data on Flow Resistance and Heat Transfer for Cross Flow of Gases Over Tube Banks," by E. D. Grimison, Trans. ASME, vol. 59, 1937, pp. 583–594.

Discussion

R. E. Dannenberg.[3] This paper is extremely interesting. It is especially useful in that the author's results give a comprehensive picture of the relationship between the pressure loss and the solidity for sharp-edged, thin perforated sheets with normal air flow.

This discusser would like to call the author's attention to additional experimental results, published subsequent to (17),[4] in NACA Technical Note 3669. In this latter note are presented the air-flow resistance characteristics of a series of perforated metal sheets having solidities ranging from 0.59 up to more than 0.99. The writer has computed the loss coefficient λ for a number of thin sheets with drilled holes (from TN 3669) and compared the values with those indicated by the author's theory in Fig. 6. The agreement was very good—within 5 per cent of the value of λ indicated from the figure for a particular solidity.

The writer feels that the actual agreement between the author's theory and experiment (for sharp-edged, straight-sided perforations) is considerably better than that implied from a comparison with some of the data in Fig. 6. Particularly in question is the effective solidity of the punched sheets used in (16, 17, and 18). With punched sheets, the punching process generally leaves small burrs or tends to dimple the metal around a perforation. No attempt is made to file the edges of the holes smooth. The writer has found that the air-flow resistance characteristics of punched sheets with the same geometric pattern, but in different metals, differed by an amount equivalent to a solidity change as large as 0.06. Reversing the direction of air flow through a sheet had no appreciable effect.

One application of sharp-edged perforated screens, which is very much alive today, is as a porous material for area-suction boundary-layer control. Aerodynamic systems include such diverse installations as supersonic inlets, high-lift devices, and wind-tunnel walls. Research has indicated that the solidity range of interest is from about 0.90 to 0.98. Although inclined or parallel flow is quite likely to occur in aerodynamic installations, knowl-

edge of the normal flow calibration is always desirable. It would appear to be very useful to extend the theoretical curves in Figs. 6 and 8 to a solidity of 0.98 or 0.99. Some experimental data in this range are available in the aforementioned NACA paper. It would be interesting to have the author's comments on the present theory in the very high solidity range.

Elmo G. Peterson.[5] In the investigation of losses in flow through screens, one must consider the basic reasons for these losses. Four types of losses are covered in the present investigations, i.e. (1) laminar flow losses caused by boundary layer shearing stresses, (2) turbulence generation losses caused by conversion of potential energy to random turbulence which degenerates because of shearing forces into heat, (3) compressibility losses caused by transfer of kinetic energy to heat ahead of the screen or immediately downstream, and (4) flow acceleration losses occurring at the constriction.

All four of these losses are involved in flow through round wire screens. In the present paper all four have been lumped into one value and equations fitted to the experimental data where possible. In turbulent flow between Reynolds numbers of 10^3 and 10^4 the generation of turbulence between jets and wakes causes the major loss. It is here that correlation of data between round wire and square bar or flat plate screens should depend mainly upon the effective solidity ratio; i.e., the minimum jet area divided by the total area. When more turbulence data are available, it may be possible to develop equations for predicting this type of energy transfer. It is interesting to note that using the maximum values of v/V given in fig. 9 of reference (19) and squaring them (to get an energy value equivalent to $\Delta p/\frac{1}{2}\rho V_0^2$) and multiplying by 60 yields values of the same magnitude as the pressure drop.

Future investigators of flow through screens should study also one of the factors affecting the pressure drop. Relative intensity of turbulence and the jet shape seem to be the easiest to study at the present time. Comparisons of these factors between low and high Mach numbers would also yield information about the compressibility effects and lead to development of an equation involving the turbulence losses as a function of s, screen form, and perhaps **M** and compressibility losses as a function of the same variables. The values obtained from Fig. 4 yield the following equation

$$\frac{\Delta p_{(v_0=s)} - \Delta p_{(\mathbf{M}_1=0)}}{\frac{1}{2}\rho V_0^2 \left(\frac{1}{1-s}\right)^2} \cong 0.3$$

The author's correlation of data and theory for round-bar screens is a major step toward development of usable equations, and the curves presented for energy losses under all conditions should satisfy most users of screens.

A. N. Thomas, Jr.[6] This paper is of considerable interest to those of us engaged in the development of air breathing powerplants, and is a valuable addition to a field in which work of this type is badly needed; particularly in regard to the reduction of large scale nonuniformities in engine inlet ducts. The distribution of flow encountered by a turbojet compressor influences the pressure ratio and consequently the over-all performance of the engine. Flow distribution is also of considerable importance to the performance of a ramjet engine. A uniform distribution of fuel and air flow widens the stability limits and increases the combus-

tion efficiency. The stronger shock wave-boundary layer interaction effects at increased Mach numbers and the higher angles of attack required at high altitude increase the difficulty of maintaining a uniform flow distribution. Therefore the use of screens or grids for the reduction of large scale velocity and pressure nonuniformities is of considerable importance. However, the interest in these devices depends upon their capacity to modify flow distribution compared to the total pressure losses incurred.

This discusser has had some experience with the reduction of large scale nonuniformities that indicates the redistribution capacity of a grid or screen is uniquely related to the Mach number reached in the minimum area. The closer the device is designed to choke the flow (increased duct blockage) the greater the capacity to reduce large scale deviations in the flow distribution. The implication appears to be that choking occurs in the screen or grid in sections of the duct carrying an excess of flow. This choking forces a crossover of the excess air into the low flow rate regions.

Unfortunately the use of a high blockage screen of the sharp-edged type analyzed by the author, will lead to prohibitive total pressure losses for the air breathing powerplant application. A similar conclusion has been reached for conventional screens of the round-wire type.

Apparently these high total pressure losses can be considerably alleviated by the use of screens or grids composed of streamlined elements. Experience of the discusser indicates that a fourfold reduction in the total pressure increment can be accomplished by the use of low drag grid elements without jeopardizing the capacity of the device to reduce large scale flow nonuniformities. However, additional analysis of the type contributed by the author is needed in this area before a complete understanding of the flow process involved can be accomplished.

Long Sun Tong.[7] The writer wishes to congratulate the author for the good agreement between the experimental data and author's theoretical prediction of the loss coefficient of sharp-edged screens.

It is also very interesting to compare the predicted loss coefficient rise of round-wire screen due to compressibility with the experimental results.

From Fig. 4, for $s = 0.6$, $\mathbf{M}_1$ increases from 0.1 to 0.18, the percentage rise of λ is predicted as

$$\frac{\lambda(\mathbf{M}_1) - \lambda(0)}{\lambda(0)} = \frac{2.75 - 2.3}{2.3} = 0.2 \text{ or } 20\%$$

From Table 2, screen No. 1, $s = 0.665$,

$$\mathbf{M}_1 = 0.1009, \quad \lambda = 4.82$$
$$\mathbf{M}_1 = 0.1834, \quad \lambda = 5.83$$

the percentage rise of λ is measured as

$$\frac{\lambda(0.18) - \lambda(0.1)}{\lambda(0.1)} = \frac{5.83 - 4.82}{4.82} = 0.21 \text{ or } 21\%$$

This good agreement shows that the curve rise at Re = 17,000 in Fig. 3 is mainly due to compressibility effect rather than Reynolds number effect. In order to investigate pure Reynolds number effect around Re = 17,000, a screen of large wire diameter (0.3 in. or up) would be good for use in air tests provided that the testing equipment is suitable for a big duct.

Author's Closure

The author is grateful to the discussers for their interesting and valuable comments. Mr. Dannenberg's new data on sharp

edged screens[8] are useful in extending the range of the low velocity empirical data of Fig. 6 up to $s = 0.995$. The theoretical curve may be extended by taking $\varphi \sim 0.611$, Kirchoff's theoretical (13) value for $s = 1$, the case of a single opening in an infinite plate. With $\varphi \sim 0.611$, Equation [3] yields a simple expression for λ, valid for $s > 0.90$, as can be demonstrated by comparison with the theoretical curves of Fig. 6. Study of Mr. Dannenberg's loss data for punched and drilled plates shows that the former ($0.944 < s < 0.994$) agree more closely with theory than the data of Czarnecki (16) shown in Fig. 6, but still are consistently lower than theory would predict. The drilled plate data ($0.923 < s < 0.995$) agree still more closely with theory, scattering above and below the curve, but less closely than the 5 per cent noted by Mr. Dannenberg for lower solidity data. In order to extend the theoretical compressible flow curves of Fig. 8 to higher solidity, additional data are required for contraction coefficient, $\varphi(\mathbf{M}_1, s)$ at higher s. The theoretical values of Chaplygin[9] for a single opening in an infinite plate might be used.

Mr. Peterson's approximate relation giving $\lambda(1) - \lambda(0)$ in

terms of s is useful in estimating choking loss coefficient from incompressible data on round-wire screens. It is indeed true that much more work is needed to clarify the roles of the various phenomena contributing to round-wire screen losses.

Screens composed of streamlined elements of the type discussed by Mr. Thomas have been used successfully for protecting jet engine inlets from foreign objects, without introducing large pressure losses. The writer is familiar with relatively little generalized loss data for such screens, especially in the compressible flow regime.

The author agrees with Dr. Tong that compressibility effect causes the increase of loss with Reynolds number for each of the two sets of experimental data shown in Fig. 3 for the present round-wire screens. More high Reynolds number, low Mach number data are required to justify the present contention regarding the high Reynolds number data. At least, it seems evident that loss factor $\lambda(1 - s)^2/s$ does not fall continuously as Reynolds number increases in the range $400 < Re < 17,000$, but rather the curve at least flattens out and probably rises somewhat.

[8] "Perforated Sheets as a Porous Material for Distributed Suction and Injection," by R. E. Dannenberg, B. J. Gambucci, and J. A. Weiberg. NACA TN 3669, 1956.

[9] "On Gas Jets," by S. A. Chaplygin, Sci. Annals Imp. Univ. Moscow, Phys.-Math. Div. 21, Moscow, 1904. Also NACA TM 1063, 1944.

The Effect of Heat Conductance on Slider-Bearing Characteristics[1]

BY W. H. GUILINGER[2] AND E. A. SAIBEL,[3] PITTSBURGH, PA.

The slider bearing without side leakage is considered when heat generated in the oil film may be conducted in the fluid as well as lost to the surroundings, which are assumed held at fixed temperatures. The basic equations are reduced to a form suitable for calculation using a digital computer, and a procedure for carrying out these calculations is suggested together with error criteria.

Introduction

SOLUTIONS to the problem of the pressure distribution, as well as other properties of interest, of the slider bearing have in the past mostly concerned themselves with isothermal behavior. Some work has been done on the assumption of adiabatic oil flow in which the viscosity of the oil is considered a function of temperature and pressure (5).[4]

However, very little is known when the real situation occurs, that is, when the heat generated in the oil is conducted in the oil and through the bearing and slider surfaces. The reason is, of course, that the resulting equations become quite intractable.

With the advent of the digital computer and its widespread availability, it is now possible to solve these problems. It is the object of the present paper to consider the problem of the slider bearing when the heat generated in the oil by its internal friction affects the density and viscosity, and when the pressure generated in the oil affects the density and viscosity, and when this heat may be conducted in the fluid and through the bearing and slider surfaces.

Starting from the Navier-Stokes equations, the equation of continuity, and the equation of dissipation, by considering orders of magnitude as is generally done in boundary-layer theory, these equations are reduced to tractable form.

The equations are then put in form for machine calculation and a method of carrying out the calculations as well as a type of error control are suggested.

Basic Equations

The following equations are valid inside the oil film between the slider and bearing as depicted in Fig. 1, for $0 \leq x \leq B$ and $0 \leq z \leq h(x)$. The bearing is stationary and the slider moves in the x-direction with speed U. The motion is assumed to be independent of time. Also, the bearing and slider are assumed to be rigid, without deformations due to changes in pressure or tem-
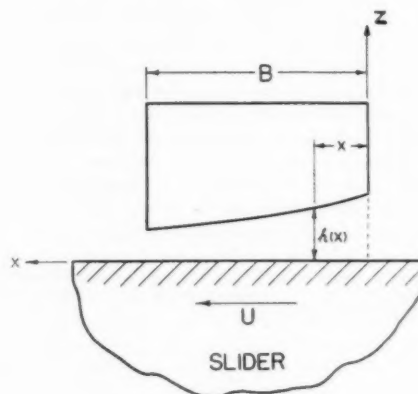
Fig. 1

perature. Other assumptions will be stated in the course of the development of the equations used. Derivation of the basic equations may be found in reference (1), among others.

The continuity equation is

$$\frac{\partial}{\partial x}\left[\rho(x, z)u(x, z)\right] + \frac{\partial}{\partial z}\left[\rho(x, z)w(x, z)\right] = 0 \dots\dots [1]$$

where ρ is the density of the oil, u is the horizontal component, and w is the vertical component of velocity of the oil.

Newton's law of motion in the x-direction yields

$$\rho(x, z)\left[u(x, z)\frac{\partial u(x, z)}{\partial x} + w(x, z)\frac{\partial u(x, z)}{\partial z}\right]$$

$$= -\frac{\partial p(x, z)}{\partial x} + \frac{\partial}{\partial x}\left[\mu(x, z)\frac{\partial u(x, z)}{\partial x}\right]$$

$$+ \frac{\partial}{\partial z}\left[\mu(x, z)\frac{\partial u(x, z)}{\partial z}\right] \dots\dots [2]$$

where p is the pressure and μ is the viscosity of the oil.

Newton's law of motion in the z-direction similarly yields

$$\rho(x, z)\left[u(x, z)\frac{\partial w(x, z)}{\partial x} + w(x, z)\frac{\partial w(x, z)}{\partial z}\right]$$

$$= -\frac{\partial p(x, z)}{\partial z} + \frac{\partial}{\partial x}\left[\mu(x, z)\frac{\partial w(x, z)}{\partial x}\right]$$

$$+ \frac{\partial}{\partial z}\left[\mu(x, z)\frac{\partial w(x, z)}{\partial z}\right] \dots\dots [3]$$

The energy equation is

$$c\rho(x, z)\left[u(x, z)\frac{\partial T(x, z)}{\partial x} + w(x, z)\frac{\partial T(x, z)}{\partial z}\right]$$

$$= k\left[\frac{\partial^2 T(x, z)}{\partial x^2} + \frac{\partial^2 T(x, z)}{\partial z^2}\right] + \mu(x, z)\Phi(x, z) \dots\dots [4]$$

where T is the oil temperature, c is the specific heat, k is the heat-conduction constant (thermal conductivity), and Φ is the dissipation function given by, terms of a smaller order of magnitude being neglected[5]

$$\Phi(x, z) = 2\left[\frac{\partial u(x, z)}{\partial x}\right]^2 + 2\left[\frac{\partial w(x, z)}{\partial z}\right]^2$$
$$+ \left[\frac{\partial w(x, z)}{\partial x} + \frac{\partial u(x, z)}{\partial z}\right]^2 \ldots \ldots [5]$$

Boundary Conditions and Further Assumptions

The assumptions that we shall make are exactly those used in the development of the well-known Reynolds equation (2), except that we shall allow the temperature to vary in the z-direction. Also, the bearing and slider will be assumed to be held at fixed temperatures, with the flow of heat between the oil film and the bearing and slider governed by Newton's law of cooling (1).

We shall assume that $h(x)$ and $dh(x)/(dx)$ are small. With this assumption and the assumption that the flow is laminar instead of turbulent (2), it is reasonable to neglect the inertia terms in Equations [2] and [3]; that is, the left-hand sides of those equations will be considered zero. Although the density ρ is then deleted from those equations, the effect of ρ varying with temperature and pressure is still retained in Equations [1] and [4].

Another assumption that we make is that w, the vertical component of velocity of the oil film, is negligible and will be assumed zero.

Under the foregoing assumptions, Equations [1] through [5] become

$$\frac{\partial}{\partial x}[\rho(x, z)u(x, z)] = 0 \ldots \ldots \ldots [6]$$

$$\frac{\partial p(x, z)}{\partial x} = \frac{\partial}{\partial x}\left[\mu(x, z)\frac{\partial u(x, z)}{\partial x}\right]$$
$$+ \frac{\partial}{\partial z}\left[\mu(x, z)\frac{\partial u(x, z)}{\partial z}\right] \ldots \ldots [7]$$

$$\frac{\partial p(x, z)}{\partial z} = 0 \ldots \ldots \ldots [8]$$

$$c\rho(x, z)u(x, z)\frac{\partial T(x, z)}{\partial x}$$
$$= k\left[\frac{\partial^2 T(x, z)}{\partial x^2} + \frac{\partial^2 T(x, z)}{\partial z^2}\right] + \mu(x, z)\Phi(x, z) \ldots \ldots [9]$$

$$\Phi(x, z) = 2\left[\frac{\partial u(x, z)}{\partial x}\right]^2 + \left[\frac{\partial u(x, z)}{\partial z}\right]^2 \ldots \ldots [10]$$

In the interest of mathematical simplicity, we shall make the usual assumption, that, in Equation [7] the variation of u with respect to x can be neglected in comparison with the z-variation of u. Then Equation [7] becomes

$$\frac{\partial p(x, z)}{\partial x} = \frac{\partial}{\partial z}\left[\mu(x, z)\frac{\partial u(x, z)}{\partial z}\right] \ldots \ldots [11]$$

Equation [8] is satisfied only if the pressure is independent of z. Thus integrating Equation [11] from zero to z we obtain

$$\frac{\partial u(x, z)}{\partial z} = \frac{z}{\mu(x, z)}\frac{dp(x)}{dx} + \frac{A(x)}{\mu(x, z)} \ldots \ldots [12]$$

Integrating once more we obtain

[5] Reference (1), p. 259.

$$u(x, z) = \frac{dp(x)}{dx}\int_0^z\frac{ydy}{\mu(x, y)} + A(x)\int_0^z\frac{dy}{\mu(x, y)} + B(x) \ldots [13]$$

The boundary conditions on u are

$$u(x, 0) = U \ldots \ldots \ldots [14]$$

and

$$u(x, h) = 0 \ldots \ldots \ldots [15]$$

Solving Equations [13], [14], and [15] for $A(x)$ and $B(x)$, we obtain

$$A(x) = -\frac{U + \frac{dp(x)}{dx}\int_0^{h(x)}\frac{zdz}{\mu(x, z)}}{\int_0^{h(x)}\frac{dz}{\mu(x, z)}} \ldots \ldots \ldots [16]$$

and

$$B(x) = U \ldots \ldots \ldots [17]$$

Let us define $r(x)$, such that

$$r(x) = \int_0^{h(x)}\rho(x, z)u(x, z)dz \ldots \ldots \ldots [18]$$

Then

$$\frac{dr(x)}{dx} = \int_0^{h(x)}\frac{\partial}{\partial x}[\rho(x, z)u(x, z)]dz$$
$$+ \rho(x, h)u(x, h)\frac{dh(x)}{dx} \ldots \ldots [19]$$

The first term on the right side of Equation [19] vanishes because of Equation [6], and the second term on the right side vanishes because of Equation [15]. Thus we obtain

$$\frac{dr(x)}{dx} = 0 \ldots \ldots \ldots [20]$$

Multiplying both sides of Equation [13] by $\rho(x, z)$ and integrating from zero to $h(x)$ we obtain

$$r(x) = \frac{dp(x)}{dx}\int_0^{h(x)}\rho(x, z)dz\int_0^z\frac{ydy}{\mu(x, y)}$$
$$+ A(x)\int_0^{h(x)}\rho(x, z)dz\int_0^z\frac{dy}{\mu(x, y)} + U\int_0^{h(x)}\rho(x, z)dz$$
$$\ldots \ldots [21]$$

If ρ and μ are independent of z, then with

$$q_x = \frac{r(x)}{\rho(x)} \ldots \ldots \ldots [22]$$

Equations [20] and [21] reduce to

$$\frac{d}{dx}[\rho q_x] = 0 \ldots \ldots \ldots [23]$$

and

$$q_x = \frac{Uh}{2} - \frac{h^3}{12\mu}\frac{dp}{dx} \ldots \ldots \ldots [24]$$

from which one obtains the well-known Reynolds equation.

Equations [20] and [21] will be used in the development of the pressure solution.

The boundary conditions that we shall assume for the pressure are

$$p(0) = 0 \ldots \ldots \ldots [25]$$

and
$$p(B) = 0 \quad\quad\quad [26]$$

We shall assume that the bearing and slider are held at fixed temperatures, T_b and T_s, respectively. Also, we shall assume that the entrance oil film temperature is some fixed T_e, independent of z, and that the exit oil film temperature ($x = B$) satisfies

$$\frac{\partial T(x, z)}{\partial x} = 0 \quad\quad\quad [27]$$

for all values of z. As was previously mentioned, we shall assume that heat flows between the oil film and the bearing and slider according to Newton's law of cooling.

Actually, the solution of Equation [9] under the assumptions outlined is a two-dimensional problem which could be solved. However, instead of solving it as a two-dimensional problem, we are going to make an approximation that will reduce the solution of Equation [9] to a one-dimensional problem. This can be solved much more readily than the two-dimensional problem. The approximation we make is that

$$T(x, z) = \xi(x) + z\eta(x) + z^2\zeta(x) \quad\quad\quad [28]$$

where ξ, η, and ζ are unknown functions of x.

Then from Newton's law of cooling we obtain

$$\left.\frac{\partial T(x, z)}{\partial z}\right|_{z = h(x)} = H_b[T_b^{\bullet} - T(x, h)] \quad\quad\quad [29]$$

and

$$-\left.\frac{\partial T(x, z)}{\partial z}\right|_{z=0} = H_s[T_s - T(x, 0)] \quad\quad\quad [30]$$

where H_b and H_s are the rates of conduction of the bearing and slider, respectively.

Eliminating $\eta(x)$ and $\zeta(x)$ from Equation [28] by solving for them in Equations [29] and [30], we obtain

$$T(x, z) = \left[1 + zH_s - \frac{z^2(H_b + H_s + hH_bH_s)}{h(2 + hH_b)}\right]\xi(x)$$
$$+ \left[\frac{z^2}{h(2 + hH_b)}\right]H_bT_b + \left[-z + \frac{z^2(1 + hH_b)}{h(2 + hH_b)}\right]H_sT,$$
$$\quad\quad\quad [31]$$

Integrating Equation [9] from zero to $h(x)$ and using Equations [29] and [30] we obtain

$$c\int_0^{h(x)} \rho(x, z)u(x, z)\frac{\partial T(x, z)}{\partial x}dz = k\int_0^{h(x)}\frac{\partial^2 T(x, z)}{\partial x^2}dz$$
$$+ kH_b[T_b - T(x, h)] + kH_s[T_s - T(x, 0)]$$
$$+ \int_0^{h(x)}\mu(x, z)\Phi(x, z)dz \quad\quad\quad [32]$$

Equations [31] and [32] will be used in the development of the temperature solution.

Since we are approximating the z-variation of the temperature distribution by Equation [28], the entrance and exit boundary conditions on the temperature cannot be satisfied for all values of z. In view of the foregoing, we shall use the conditions on the average of T across the thickness of the film; namely

$$\frac{1}{h(0)}\int_0^{h(0)} T(0, z)dz = T_e \quad\quad\quad [33]$$

and

$$\int_0^{h(B)}\left.\frac{\partial T(x, z)}{\partial x}\right|_{x = B} dz = 0 \quad\quad\quad [34]$$

Pressure Solution

Let $0 \leq z \leq B$ be broken up into N equal intervals of length B/N. Let $x_n = \frac{n}{N}B$ for $n = 0, 1, \ldots, N$. Also, let $\rho(x_n, z) = \rho_n(z)$, etc.

Then Equation [21] can be rewritten as

$$r_n = \left.\frac{dp(x)}{dx}\right|_{x=x_n}\int_0^{h_n}\rho_n(z)dz\int_0^z\frac{ydy}{\mu_n(y)}$$
$$+ A_n\int_0^{h_n}\rho_n(z)dz\int_0^z\frac{dy}{\mu_n(y)} + U\int_0^{h_n}\rho_n(z)dz \quad\quad\quad [35]$$

Equation [20] and $\left.\frac{dp}{dx}\right|_{x=x_n}$ can be approximated by (3)

$$r_{n+1} - r_n = 0 \quad\quad\quad [36]$$

and

$$\left.\frac{dp(x)}{dx}\right|_{x=x_n} = \frac{p_n - p_{n-1}}{\frac{B}{N}} \quad\quad\quad [37]$$

for $n = 1, 2, \ldots, N - 1$. Substituting Equations [37] and [35] into [36] we obtain

$$a_n p_{n-1} - (a_n + a_{n-1})p_n + a_{n-1}p_{n-1} = b_n - b_{n-1} \quad\quad\quad [38]$$

for $n = 1, 2, \ldots, N - 1$ where

$$a_n = \int_0^{h_{n+1}}\rho_{n+1}(z)dz\int_0^z\frac{ydy}{\mu_{n+1}(y)}$$
$$- \frac{\int_0^{h_{n+1}}\frac{zdz}{\mu_{n+1}(z)}}{\int_0^{h_{n+1}}\frac{dz}{\mu_{n+1}(z)}}\int_0^{h_{n+1}}\rho_{n+1}(z)dz\int_0^z\frac{dy}{\mu_{n+1}(y)} \quad\quad\quad [39]$$

and

$$b_n = \frac{B}{N}U\frac{\int_0^{h_{n+1}}\rho_{n+1}(z)dz\int_0^z\frac{dy}{\mu_{n+1}(y)}}{\int_0^{h_{n+1}}\frac{dz}{\mu_{n+1}(z)}}$$
$$- \frac{B}{N}U\int_0^{h_{n+1}}\rho_{n+1}(z)dz \quad\quad\quad [40]$$

Let $0 \leq z \leq h_n$ be broken up into M equal intervals of length h_n/M. Let $z_{nm} = \frac{m}{M}h_n$ for $m = 0, 1, \ldots, M$. Then the integral from 0 to z_{nm} of an arbitrary function $f(z)$ can be approximated by (3)

$$\int_0^{z_{nm}} f(z)dz = \sum_{i=0}^m \alpha_{ni}{}^m f(z_{ni}) \quad\quad\quad [41]$$

For example, if the trapezoidal approximation is used then

$$\alpha_{ni}{}^m = \frac{h_n}{M} \quad \text{for} \quad i = 1, 2, \ldots, m - 1 \quad\quad\quad [42]$$

and

$$\alpha_{ni}{}^m = \frac{h_n}{2M} \quad \text{for} \quad i = 0 \text{ and } i = m \quad\quad\quad [43]$$

Let us suppose that we have a guess of p_n and ξ_n say $p_n{}^{(0)}$ and

$\hat{\xi}_n$ for $n = 1, 2, \ldots, N$. Then by using Equation [41] we can calculate $a_n^{(0)}$ and $b_n^{(0)}$ for $n = 0, 1, \ldots, N - 1$. Let the solution of Equation [38] then be labeled $p_n^{(1)}$. Thus we want to solve the system

$$a_n^{(0)}p_{n+1}^{(1)} - [a_n^{(0)} + a_{n-1}^{(0)}]p_n^{(1)} + a_{n-1}^{(0)}p_{n-1}^{(1)}$$
$$= b_n^{(0)} - b_{n-1}^{(0)} \ldots \ldots [44]$$

Let us assume for the moment that it is possible to solve Equation [44] for $n = 1, 2, \ldots, N - 1$. Then with $p_n^{(1)}$ we can calculate $a_n^{(1)}$ and $b_n^{(1)}$ and then solve Equation [41] for $p_n^{(2)}$. After i such iterations, we would like to solve the system

$$a_n^{(i)}p_{n+1}^{(i+1)} - [a_n^{(i)} + a_{n-1}^{(i)}]p_n^{(i+1)}$$
$$+ a_{n-1}^{(i)}p_{n-1}^{(i+1)} = b_n^{(i)} - b_{n-1}^{(i)} \ldots \ldots [45]$$

for $n = 1, 2, \ldots, N - 1$. Equation [45] with the Boundary Conditions [25] and [26] is in just the right form for solution by the three-term linear system method (4).

Consider

$$\alpha_n^{(i)} = \frac{b_n^{(i)} - b_{n-1}^{(i)} - \alpha_{n-1}^{(i)}a_{n-1}^{(i)}}{\beta_{n-1}^{(i)}a_{n-1}^{(i)} - a_n^{(i)} - a_{n-1}^{(i)}} \ldots \ldots [46]$$

and

$$\beta_n^{(i)} = \frac{-a_n^{(i)}}{\beta_{n-1}^{(i)}a_{n-1}^{(i)} - a_n^{(i)} - a_{n-1}^{(i)}} \ldots \ldots [47]$$

to be calculated for $n = 1, 2, \ldots, N - 1$ with $\alpha_0^{(i)} = \beta_0^{(i)} = 0$. Then it can be shown that

$$p_{N-1}^{(i+1)} = \alpha_{N-1}^{(i)} \ldots \ldots \ldots [48]$$

and

$$p_n^{(i+1)} = \alpha_n^{(i)} + \beta_n^{(i)}p_{n+1}^{(i+1)} \ldots \ldots [49]$$

for $n = N - 2, N - 3, \ldots, 2, 1$.

It is not known *a priori* that the foregoing iterative process converges. Assuming that it does, we need some error criterion to terminate the process. One possibility is

$$\underset{n = 1, \ldots, N-1}{\text{Max}} \left| p_n^{(i+1)} - p_n^{(i)} \right| < \epsilon_1 \ldots \ldots [50]$$

where ϵ_1 is some prescribed positive number. Thus if the iterative process converges, given a guess of the pressure and temperature distributions we can obtain a new pressure distribution which, with the old temperature distribution, satisfies Equation [38].

Temperature Solution

Substituting the approximations

$$\left. \frac{\partial T(x, z)}{\partial x} \right|_{x = x_n} = \frac{N}{2B} [T_{n+1}(z) - T_{n-1}(z)] \ldots \ldots [51]$$

and

$$\left. \frac{\partial^2 T(x, z)}{\partial x^2} \right|_{x = x_n} = \left(\frac{N}{B} \right)^2 [T_{n+1}(z) - 2T_n(z) + T_{n-1}(z)] \ldots [52]$$

into Equation [32] we obtain

$$\frac{NC}{2B} \int_0^{h_n} \rho_n(z)u_n(z)[T_{n+1}(z) - T_{n-1}(z)]dz$$
$$= \left(\frac{N}{B} \right)^2 k \int_0^{h_n} [T_{n+1}(z) - 2T_n(z) + T_{n-1}(z)]dz$$
$$+ kH_b[T_b - T_n(h_n)] + kH_s[T_s - T_n(0)]$$
$$+ \int_0^{h_n} \mu_n(z)\Phi_n(z)dz \ldots \ldots [53]$$

Substituting the approximations

$$\left. \frac{dp(x)}{dx} \right|_{x = x_n} = \frac{N}{2B} [p_{n+1} - p_{n-1}] \ldots \ldots [54]$$

and

$$\left. \frac{\partial u(x, z)}{\partial x} \right|_{x = x_n} = \frac{N}{2B} [u_{n+1}(z) - u_{n-1}(z)] \ldots \ldots [55]$$

into Equations [12] and [10], respectively, we obtain

$$\Phi_n(z) = \frac{1}{2} \left(\frac{N}{B} \right)^2 [u_{n+1}(z) - u_{n-1}(z)]^2$$
$$+ \left\{ \frac{Nz}{2B\mu_n(z)} [p_{n+1} - p_{n-1}] - \frac{U + [p_{n+1} - p_{n-1}]\frac{N}{2B}\int_0^{h_n} \frac{zdz}{\mu_n(z)}}{\mu_n(z)\int_0^{h_n} \frac{dz}{\mu_n(z)}} \right\}^2$$
$$\ldots \ldots [56]$$

Also, substituting Equation [54] into [16] and [13], we obtain

$$u_n(z) = U + \frac{N}{2B} [p_{n+1} - p_{n-1}] \int_0^z \frac{ydy}{\mu_n(y)}$$
$$- \frac{U + \frac{N}{2B} [p_{n+1} - p_{n-1}] \int_0^{h_n} \frac{zdz}{\mu_n(z)}}{\int_0^{h_n} \frac{dz}{\mu_n(z)}} \int_0^z \frac{dy}{\mu_n(y)} \ldots \ldots [57]$$

Substituting Equations [31], [56], and [57] into [53], we obtain

$$d_n\xi_{n+1} + e_n\xi_n + t_n\xi_{n-1} = g_n \ldots \ldots [58]$$

for $n = 1, 2, \ldots, N - 1$ where with

$$\varphi_n(z) = 1 + \frac{zH_s}{2} - \frac{z^2}{3h_n} \left[\frac{H_b + H_s + h_nH_bH_s}{2 + h_nH_b} \right] \ldots [59]$$

and

$$\psi_n(z) = 1 + zH_s - \frac{z^2}{h_n} \left[\frac{H_b + H_s + h_nH_bH_s}{2 + h_nH_b} \right] \ldots [60]$$

we obtain

$$d_n = \frac{NC}{2B} \int_0^{h_n} \rho_n(z)u_n(z)\psi_{n+1}(z)dz$$
$$- k h_n \left(\frac{N}{B} \right)^2 \varphi_{n+1}(h_n) \ldots \ldots [61]$$

$$e_n = kH_s + 2kh_n \left(\frac{N}{B} \right)^2 \varphi_n(h_n) + kH_b\psi_n(h_n) \ldots \ldots [62]$$

$$t_n = -\frac{NC}{2B} \int_0^{h_n} \rho_n(z)u_n(z)\psi_{n-1}(z)dz$$
$$- k h_n \left(\frac{N}{B} \right)^2 \varphi_{n-1}(h_n) \ldots \ldots [63]$$

Also, with

$$\gamma_n = \frac{H_bT_b + H_sT_s + h_nH_bH_sT_s}{h_n(2 + h_nH_b)} \ldots \ldots [64]$$

we obtain

$$g_n = kH_bT_b + kH_sT_s + \int_0^{h_n} \mu_n(z)\Phi_n(z)dz$$

$$- \frac{NC}{2B} [\gamma_{n+1} - \gamma_{n-1}] \int_0^{h_n} \rho_n(z) u_n(z) z^2 dz$$

$$+ \frac{k}{3} \left(\frac{Nh_n}{B}\right)^3 [\gamma_{n+1} - 2\gamma_n + \gamma_{n-1}]$$

$$- kh_n H_b[\gamma_n h_n - H_s T_s] \ldots \ldots [65]$$

Substituting Equation [31] into [33] we obtain

$$\xi_0 = \frac{1}{\varphi_0(h_0)} \left[T_s + \frac{h_0 H_s T_s}{2} - \frac{\gamma_0 h_0}{3} \right] \ldots \ldots [66]$$

Substituting the approximation

$$\left. \frac{\partial T(x, z)}{\partial x} \right|_{x=x_n} = \frac{N}{B} [T_n(z) - T_{n-1}(z)] \ldots \ldots [67]$$

and Equation [31] into [34] we obtain

$$\xi_N = \frac{\varphi_{N-1}(h_N)}{\varphi_N(h_N)} \xi_{N-1} + \frac{h_N^2}{3\varphi_N(h_N)} [\gamma_{N-1} - \gamma_N] \ldots \ldots [68]$$

The solution of Equation [58] with the conditions [66] and [68] will be quite similar to the pressure solution of Equation [38] with conditions [25] and [26].

First, let us assume a guess of p_n and ξ_n, say p_n and $\xi_n^{(0)}$ for $n = 1, 2, \ldots, N$. Then we can calculate $d_n^{(0)}$, $e_n^{(0)}$, $f_n^{(0)}$, and $g_n^{(0)}$ for $n = 1, 2, \ldots, N-1$. Let the solution of [58] then be labeled $\xi_n^{(1)}$. Thus we want to solve the system

$$d_n^{(0)}\xi_{n+1}^{(1)} + e_n^{(0)}\xi_n^{(1)} + f_n^{(0)}\xi_{n-1}^{(1)} = g_n^{(0)} \ldots \ldots [69]$$

with [66] and [68] for $n = 1, 2, \ldots, N-1$. With $\xi_n^{(1)}$ we can calculate $d_n^{(1)}$, $f_n^{(1)}$, and $g_n^{(1)}$ and solve [58] for $\xi_n^{(2)}$. After j such iterations we would like to solve the system

$$d_n^{(j)}\xi_{n+1}^{(j+1)} + e_n^{(j)}\xi_n^{(j+1)} + f_n^{(j)}\xi_{n-1}^{(j+1)} = g_n^{(j)} \ldots [70]$$

with Equations [66] and [68] for $n = 1, 2, \ldots, N-1$.

Equation [70] with the boundary conditions [66] and [68] is in just the right form for solution by the three-term linear system method (4).

Consider

$$\pi_n^{(j)} = \frac{g_n^{(j)} - \pi_{n-1}^{(j)}f_n^{(j)}}{\theta_{n-1}^{(j)}f_n^{(j)} + e_n^{(j)}} \ldots \ldots \ldots [71]$$

and

$$\theta_n^{(j)} = \frac{-d_n^{(j)}}{\theta_{n-1}^{(j)}f_n^{(j)} + e_n^{(j)}} \ldots \ldots \ldots [72]$$

to be calculated for $n = 1, 2, \ldots, N-1$ with $\theta_0^{(j)} = 0$ and $\pi_0^{(j)} = \xi_0$ obtained from [66]. With

$$D = \frac{\varphi_{N-1}(h_N)}{\varphi_N(h_N)} \ldots \ldots \ldots [73]$$

and

$$E = \frac{h_N^2}{3\varphi_N(h_N)} [\gamma_{N-1} - \gamma_N] \ldots \ldots \ldots [74]$$

from [68] and [70] we obtain

$$\xi_N^{(j+1)} = \frac{E + D\pi_{N-1}^{(j)}}{1 - D\theta_{N-1}^{(j)}} \ldots \ldots \ldots [75]$$

and

$$\xi_n^{(j+1)} = \pi_n^{(j)} + \theta_n^{(j)}\xi_{n+1}^{(j+1)} \ldots \ldots \ldots [76]$$

for $n = N - 1, N - 2, \ldots, 2, 1$.

As in the pressure solution, it is not known *a priori* that the foregoing iterative process converges. Assuming that it does, we need some error criterion to terminate the process. As in the pressure solution, let us use

$$\underset{n = 1, \ldots, N}{\text{Max}} \left| \xi_n^{(j+1)} - \xi_n^{(j)} \right| < \epsilon_2 \ldots \ldots \ldots [77]$$

where ϵ_2 is some prescribed positive number. Thus if the iterative process converges, given a guess of the pressure and temperature distributions we can obtain a new temperature distribution which, with the old pressure distribution, satisfies Equations [58] with [66] and [68].

Use may be made of some of the expressions just developed for finding other characteristics of the flow such as total load-carrying capacity W, oil flow Q, and friction force F, defined as follows

$$W = \int_0^B p(x)dx \ldots \ldots \ldots \ldots [78]$$

$$Q = \int_0^{h(0)} u(0, z)dz \ldots \ldots \ldots \ldots [79]$$

and

$$F = \int_0^B \mu(x, 0) \left. \frac{\partial u(x, z)}{\partial z} \right|_{z=0} dx \ldots \ldots \ldots [80]$$

Once the problem has been solved for the pressure and temperature, all quantities necessary for the evaluation of these integrals by numerical methods have been found at the mesh points. Naturally, one would use the same mesh for evaluation of the integrals so that all that is necessary is that he set up the particular scheme of numerical integration he wishes, and use the values at each mesh point which have already been found.

Bibliography

1 "Boundary Layer Theory," by H. Schlichting, McGraw-Hill Book Company, Inc., New York, N. Y., 1955.
2 "Hydrodynamics," by H. Lamb, Dover Publications, New York, N. Y., 1945.
3 "Introduction to Numerical Analysis," by F. B. Hilderbrand, McGraw-Hill Book Company, Inc., New York, N. Y., 1956.
4 "Multigroup Methods for Neutron Diffusion Problems," by R. Ehrlich and H. Hurwitz, Jr., *Nucleonics*, vol. 12, February, 1954, pp. 23–30.
5 "On the Solution of the Reynolds Equation for Slider-Bearing Lubrication—IV. Effect of Temperature on the Viscosity." by F. Osterle, A. Charnes, and E. Saibel, TRANS. ASME, vol. 75, 1953, pp. 1117–1123.

Discussion

J. S. Ausman.[6] Up to Equations [29] and [30] of the paper the treatment is reasonably straightforward although the authors could have neglected additional terms; namely

$$k \frac{\partial^2 T}{\partial x^2}$$

in the energy equation and $2(\partial u/\partial x)^2$ in the dissipation function. However, Equations [29] and [30] have no physical justification. The temperature gradients on the left-hand sides of the equations are proportional to the heat flux leaving the oil film, but the right-hand sides of the equations are equal to zero, since liquid lubricant in immediate contact with a wall (bearing or slider) must take on the same temperature as that wall. Hence

$$T_B - T(x, h) = 0$$

[6] Autonetics, Division of North American Aviation, Inc., Bellflower, Calif.

and

$$T_S - T(x,0) = 0$$

Apparently the authors have attempted to apply a type of equation commonly seen in heat transfer across boundary layers in which the heat flux q is expressed as

$$q = h\Delta T$$

where h is a film coefficient of heat transfer and ΔT is the temperature differential across a small but finite fluid boundary-layer thickness. In the present application no such finite layer is assumed and indeed it does not exist in the usual sense. Ordinary boundary-layer thicknesses extend far enough out from the boundary so that the inertia terms in the momentum equations (Equations [2] and [3]) are of the same order of magnitude as the viscous terms. In lubrication theory the inertia terms are generally considered to be negligible with respect to the viscous terms (this paper is no exception) which means that the entire lubrication film is an order of magnitude smaller than the usual "boundary layer."

The correct boundary conditions which should replace Equations [29] and [30] are simply

$$T(x, 0) = T_S$$

and

$$T(x, h) = T_B$$

These two conditions when applied to Equation [28] still permit elimination of two of the unknown functions of x (and do not introduce two new unknowns such as H_B and H_S) to yield

$$T(x, z) = T_S + [T_B - T_S - h^2\zeta(x)]\, \frac{z}{h} + z^2\zeta(x)$$

which should replace Equation [31]. Presumably, the solution could be carried out from this point by methods analogous to those of the authors.

F. W. Ocvirk.[7] The case of this paper, for which the authors present a mathematical method of solution, is an important one. The assumption of uniform film temperature is often made. Such an assumption is a useful first approximation which allows heat-transfer calculations to be made readily in determining the equilibrium temperature of the lubricating film. However, one wonders how good the assumption of uniform temperature is when the slider temperature T_s, the bearing temperature T'_s, and the entering film temperature T_e are widely different.

Actually, in the over-all solution of the bearing heat-transfer problem, T_s, T_b, and T_e are unknowns as well as the film temperature. Although a number of heat-transfer equations may be written for the solution of the unknowns, usually there are not sufficient equations to determine them all without making simplifying assumptions as to the relationship of the film temperature and the temperatures T_s, T_b, and T_e. It is here very much appreciated that the authors can provide a means of relating these temperatures realistically.

The authors' assumption of uniform fixed temperatures T_s and T_b for the slider and bearing surfaces are realistic in the opinion of the writer whose experience has been with journal bearings under light load and gravity oil feed. The assumption of no side leakage made by the authors is less realistic. However, it is realized that side leakage greatly complicates the mathematical problem, and the writer wishes to acknowledge the solving of a problem which is already sufficiently complex.

[7] Associate Professor, Sibley School of Mechanical Engineering, Department of Machine Design, Cornell University, Ithaca, N. Y. Mem. ASME.

It is sincerely hoped that the authors will undertake the numerical computations required to show film-temperature distributions for typical cases, and to show the effect on load capacity and friction force of such distributions.

S. E. Weidler.[8] The writer would like the authors to describe the bearing physical configuration and lubricant in which both the heat conductivity k and local surface heat-transfer coefficient H are important. One can imagine a lubricant such as NaK where the conductivity might be important. The next part of the question then follows: If H is important, how does one measure H? H, a surface local heat-transfer coefficient, is associated with a boundary layer of a few mils thick in the usual fluid-flow problem. However, in bearings, the whole fluid-flow field is only a few mils thick. Thus, the analogous region associated with the H must be at least one scale of magnitude lower than usual.

In Equation [10], the $\partial u/\partial x$ component of the dissipation function can be ignored safely in view that the order of magnitude of the ratio of the two terms, $O(R) = B^2/h^2 = 1 \times 10^6$, shows the other to be much larger. In Equations [29] and [30], if H is the usual surface heat-transfer coefficient, then there should be a k as a factor on the left-hand side. Also, at $z = h(x)$, there is a wedge-shape section in which an additional heat flux in the z-direction occurs. This heat flux contributes another term

$$-k\, \frac{\partial T(x, z)}{\partial x}\, \frac{\partial h(x)}{\partial x}$$

to Equation [29]. Thus if $T \neq f(z)$, the usual heat-transfer term

$$q = \frac{\partial}{\partial x}\left(kh\, \frac{\partial T}{\partial x}\right)$$

may be assembled. The additional term is the same order of magnitude as the first term on the right-hand side of Equation [32].

Going one step further, the two terms of the heat-conductivity term in Equation [9] may be compared and on an order-of-magnitude basis the first term of the pair can be omitted. By the same token, the additional heat term listed in the foregoing also should be ignored. Thus Equation [32] finally becomes

$$c\int_0^{h(x)} \rho(x, z)u(x, z)\, \frac{\partial T}{\partial x}\,(x, z)\, dz = H_b[T_b - T(x, h)]$$
$$+ H_s[T_s - T(x, 0)] + \int_0^{h(x)} \mu(x, z)\left[\frac{\partial u}{\partial z}\,(x, z)\right]^2 dz$$

where all the heat-parameter units are in lb, ft, sec units.

The writer would like a small exposition on the three-term linear-system method that is apparently outlined in reference (4) of the paper. In particular, he does not understand the $\beta_{n-1}^{(i)}$ factor in Equation [46] or $\theta_{n-1}^{(j)}$ factor in Equation [71]. When the central pressure of Equation [45] and the central ξ of Equation [70] are solved for and compared with Equations [49] and [76], respectively, the writer finds that β, θ-factors should be zero.

The authors apply an absolute criterion to determine when they have arrived at a solution. It is sometimes useful to use a relative criterion where the approximation is evaluated relative to the pressure at the point in question. Another technique that is especially useful in problems involving elastic distortion and where it is fairly certain that the bearing runs full, is to sub-

[8] Application Engineer, Analytical Engineering Section, General Electric Company, Schenectady, N. Y. Assoc. Mem. ASME.

stitute $p_n{}^{(i)} = 0$ where the method calculates a negative pressure. This technique is not original with the writer.

Lastly, the writer would like to see a curve containing solutions together with plots of some of the intermediate iterations.

Authors' Closure

The authors wish to thank the discussers for their stimulating and helpful comments. Mr. Ausman's discussion of the boundary conditions expressed in Equations [29] and [30] brings out the fact that actual conditions are not known and as far as the authors are aware no definitive experimental evidence is available for the selection of one set of conditions over another. The set given in Equations [29] and [30] could have been discussed somewhat more in detail in the paper and we shall do so now. The quantities H_b and H_s may be given values ranging from zero to infinity. In the former case, the implication is that there is no heat leakage from the lubricant to the metal; in the latter case the oil and the metal take on the same temperature. It is possible that the H's have an intermediate value and that a thermal boundary layer exists. One of the discussers at the San Francisco meeting suggested that the temperatures T_b and T_s in Equations [29] and [30] be replaced by T_∞, the temperature at some distance from the surface of contact of the lubricant and the metal. This, the authors feel, has some merit and should be considered. At the present time, calculations and comparisons with experimental data must be the sole guide. Eventually the problem must be looked upon from a general point of view involving the interaction of the fluid flow, the heat generation and transfer, and the elastic properties of the system as a whole.

In answer to Professor Ocvirk's remarks, we too realize that the neglect of side flow is serious, but we think that with the proper planning the finite bearing can be solved without overtaxing existing digital computing machines. The authors have made numerical calculations for the one-dimensional case for the parallel slider bearing and these results will be sent out soon for publication.

As a matter of fact, in the example chosen for calculation, as anticipated by Mr. Weidler, the local surface heat-transfer coefficient was not important in the sense that the temperature distribution in the film was not sensitive to it, but the thermal conductivity of the fluid was important. Actually it is the Peclet number which is a controlling factor. We did have in mind the possibility of lubricants with large conducting and heat transfer possibilities which is why we included terms usually omitted.

Influence of Load and Thermal Distortion on the Design of Large Thrust Bearings

By R. A. BAUDRY,[1] E. C. KUHN,[2] AND W. W. WISE[3]

High-capacity vertical waterwheel generators require large thrust bearings in which the load is carried on a thin hydrodynamic film of oil. Thermal and load distortion of the bearing parts have to be limited to a low value to permit the normal and complete formation of the oil film predicted by the hydrodynamic theory of lubrication. Excessive load and thermal distortion materially affect bearing performance and may result in bearing failure. This paper describes some design considerations given to vertical waterwheel-generator thrust bearings in order to avoid highly concentrated loading during the starting period, and to obtain minimum distortion of the pad during operation at normal speed and load.

Introduction

AN APPRECIABLE part of the available electrical energy is generated by hydroelectric units of the vertical type. The speed and size of these units are selected to obtain the most economical and efficient installation. For many years in this country the trend has been toward machines of larger rating and lower speed requiring thrust bearings of continuously increasing size and capacity.

Waterwheel generators, as other types of heavy rotating equipment, start with the large weight of the rotating parts resting on the thrust bearing. During the relatively short starting period, before an effective oil film is established, there is metal-to-metal contact between the stationary and rotating bearing surfaces. During the starting period the bearing is required to operate under adverse conditions, which necessitates the use of adequate materials, lubricants, and smooth surfaces finished to a high standard of accuracy (1).[4]

After an effective oil film has been established the bearing surfaces are no longer in contact and the bearing can operate indefinitely without any appreciable wear. The basic hydrodynamic theory of lubrication was developed many years ago (2, 3). Later developments have resulted in a better understanding of the formation of the oil film and now it is possible to predetermine the performance of a thrust bearing more accurately (4). The determination of the oil-film thickness is based on ideal assumptions, one of which is that the thrust-bearing surfaces are plane

and free of distortion. It is necessary to approach this condition in order to obtain satisfactory performance of the bearing.

On large thrust bearings, the geometry of the bearing surface is affected appreciably by distortion due to load and thermal gradients. During the starting period, such distortion may cause highly concentrated pressures which could result in localized damage to the bearing surfaces. The effect of distortion of the bearing surfaces can be aggravated by insufficient lubrication resulting from excessive aeration of the oil as discussed in a companion paper (5).

After the oil film has been established, and during normal operation, distortion of the bearing surfaces obviously must be smaller than the thickness of the oil film if metal-to-metal contact is to be avoided.

The present paper discusses load and thermal distortion of the surfaces of large thrust-bearing pads, and the necessity of supporting the pads properly to minimize distortion, thereby approaching bearing performance as predicted by theory.

Pressure Distribution on Pad During Starting Period

The unit loading on the thrust bearing of waterwheel generators during the starting period; i.e., before an oil film is established, must be limited to a value which prevents damage to the bearing surface. It is, therefore, necessary that high peak pressures be avoided and the load be distributed over the pad as uniformly as possible. On very small thrust bearings the problem of distortion during the starting period, as well as during normal operation, does not appear critical and a concentrated support under the pad is usually satisfactory. The pressure distribution on a pad with such a support can be estimated by use of the elastic-foundation theory (6, 7). Test results made with accurately machined surfaces described in a later part of this paper are in good agreement with the results of analysis by this method. Fig. 1 shows,
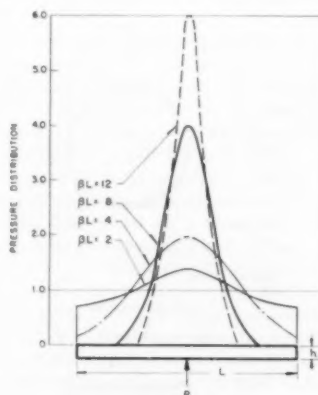

Fig. 1 Static pressure distribution on a thrust-bearing pad with a single concentrated support. (Based on calculations of a beam of unit width on an elastic foundation. $\beta = [3k/Eh^3]^{1/4}$, where k is spring constant of thrust block and runner and E is modulus of elasticity of pad material. Pressure distribution is ratio of local to average bearing pressure.)

diagrammatically, the pressure distribution on the bearing surface resulting from the use of a single concentrated support for the thrust-bearing pad. In large-diameter bearings the use of such a support is precluded in view of the need for a very thick pad to minimize peak pressures for a given elastic system. As shown by Fig. 1, a pressure in excess of four times the nominal bearing pressure is reached with relatively thin pads which are desirable from many standpoints, such as handling, babbitting, low hydrogen content, as well as the ability of the pad to conform to variations in the thrust-bearing runner surface (8).

The pressure distribution shown in Fig. 1 was based on calculation for a beam of unit width. Pads of various proportions are used and they should be considered as plates to obtain more accurate values of deflection and pressure distribution. However, the use of a beam analogy is very convenient and gives good understanding of the problem and the relations of the factors involved.

In the past, thrust bearings of the pivoted-pad type have been made with a relatively small number of pads, usually six or eight. In recent years investigations have shown it beneficial to increase the number of pads to twelve or more (9). Such pads are much longer in the radial than in the circumferential direction, for bearings of equivalent diameter, and may be considered as beams. The final analysis of a particular design, however, must be made considering the actual sector-shaped pad.

From the standpoint of bearing performance, it has been shown beneficial to have some convex distortion of the pad in the circumferential direction (10). Later unpublished work of the same authors confirms that the deflection along the radial center line of the pad is detrimental to bearing performance. Therefore, this paper gives paramount importance to minimizing the deflection of the pad along its radial center line.

High peak pressures, or the use of very thick pads, can be avoided by supporting the pads on a disk-type support, shown in Fig. 2. Fig. 3 shows the pressure distribution on a pad with this type support, as estimated by the beam on elastic-foundation theory. The disk-type support has been widely used with very good results. For larger bearings, now required owing to the continuously increasing size of low-speed hydroelectric generating units, the use of a multiple support, of the type in Fig. 4, shows that either an appreciable improvement in factor of safety or the use of higher unit pressures can be realized with a pad of the

Fig. 3 Static pressure distribution on a thrust-bearing pad resting on a disk-type support. (Based on calculation of a beam of unit width on an elastic foundation. $\beta = [3k/Eh^3]^{1/4}$, where k is spring constant of thrust block and runner and E is modulus of elasticity of pad material. Pressure distribution is ratio of local to average bearing pressure.)
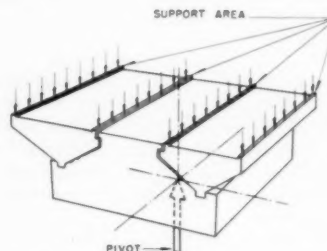
Fig. 4 A multiple support for thrust-bearing pads
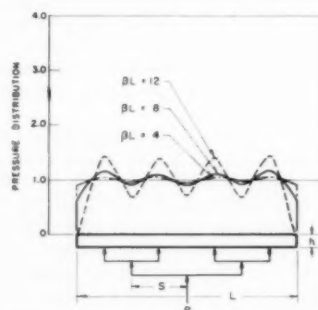
Fig. 5 Static pressure distribution on a thrust-bearing pad resting on a multiple support. (Based on calculation of a beam of unit width on an elastic foundation. $\beta = [3k/Eh^3]^{1/4}$, where k is spring constant of thrust block and runner and E is modulus of elasticity of pad material. Pressure distribution is ratio of local to average bearing pressure. Reaction location $S/L = 0.25$, optimum for static pressure distribution.)

same thickness. Fig. 5 shows, with this support, peak pressures to be less than 125 per cent of the average pressure for pad proportions now used in large thrust bearings. With the multiple support, large thrust-bearing pads can be designed with minimum thickness and possess more uniform pressure distribution during the starting period.

(a) *Measurement of Bearing Pressure Distribution During Starting Period on Sector-Shaped Thrust-Bearing Pads.* The merits of multiple supports for large thrust-bearing pads are indicated by comparison of Figs. 1, 3, and 5, based on elastic-
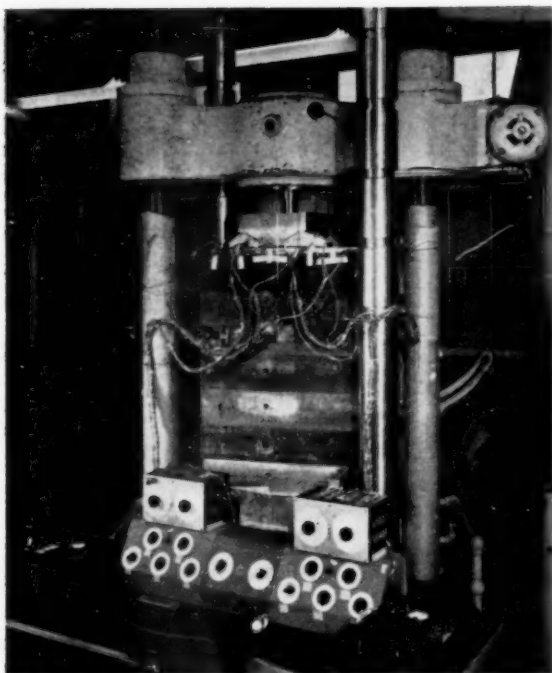
Fig. 2 A disk-type support for thrust-bearing pads

Fig. 6   Test arrangement for determining static pressure distribution on a sector-shaped pad of a 105-in-diam thrust bearing

Fig. 7   Arrangement for calibrating 116 strain gages attached to a sector-shaped pad of a 105-in-diam thrust bearing for determining static pressure distribution. (Shown, also, are $^1/_{16}$-in-wide $\times$ $^3/_8$-in-deep grooves for attaching strain gages. Fig. 22 is diagram of strain-gage locations in this sector-shaped pad.)

foundation theory.   Investigations were conducted to determine experimentally the static pressure distribution on a sector-shaped thrust-bearing pad.   The pressure distribution on a pad of a 105-in-diam bearing, when supported by a disk of the type shown in Fig. 2, also was investigated for comparison with the static pressure distribution obtained using the multiple support shown in Fig. 4.

The pad to be tested was placed in a compression testing machine in a manner duplicating the support and rigidity realized in an actual waterwheel generator.   Fig. 6 shows the arrangement of the pads in the testing machine.   As can be noticed from this view, the test pad is supported from the bottom platen by other steel pads of the same geometry.   The mating surfaces of these supporting pads were ground and scraped to intimacy and were not babbitted.   The babbitt on the test pad was bonded to the machined steel surface without any anchor grooves.   The babbitt was well bonded to the steel surface as verified by ultrasonic inspection (8).   The surface of the babbitt was prepared by scraping to a surface plate.   Such scraping resulted in a maximum variation of 0.00015 in. from a plane surface.   The babbitt on the test pad had eleven $^1/_{16}$-in-wide grooves, $^3/_8$ in. deep. Along the walls of the grooves were cemented electrical-resistance strain gages.   The pattern of strain-gage locations was such to facilitate adequate interpretation of the strains at all points on the surface of the pad.   Each gage was individually calibrated. Fig. 7 shows the test pad being subjected to the strain-gage calibration.   This view shows also, the grooves in the pad and the strain-gage leads entering the grooves.

The test pad was loaded up to 400,000 lb.   Strains at 116 points were recorded at this and several intermediate loads. Strain-gage readings were obtained using the multiple support, as well as the disk-type support.

(b)   *Results of Static Pressure Test.*   The pressures associated with these strain-gage readings were calculated from the previous
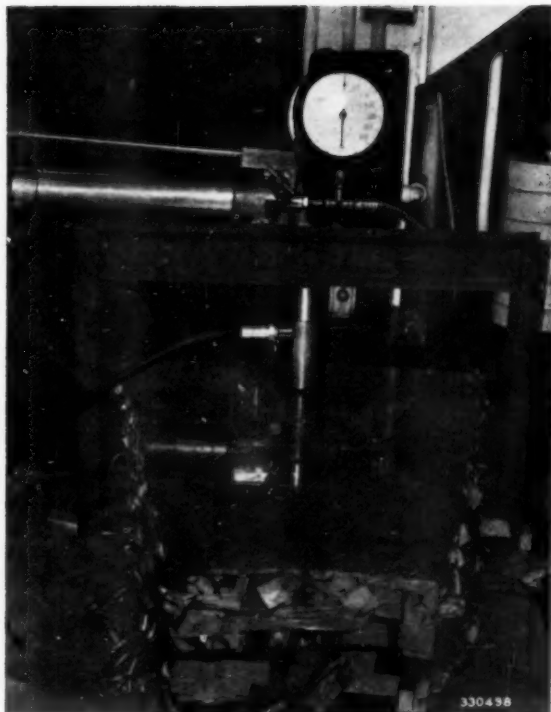
calibration.   These values, for an intermediate loading of 217,000 lb, for example, were plotted to form the contours shown in Fig. 8 for the disk-type support and in Fig. 9 for the multiple support.   For each of these plots the area between adjacent contours was determined by a planimeter and multiplied by the average value of the adjacent pressure contours.   The sum of such measurements was found in good agreement with the applied load.

For the disk-type support, as shown in Fig. 8, there are regions of high pressures corresponding to the circular outline of the disk support.   The measured load distribution is similar to that which was illustrated in Fig. 3, for a structure of equivalent rigidity.

Fig. 9 shows, for a multiple support, the pressure distribution along the radial center line of the sector-shaped pad, as well as the variation of pressure in a circumferential direction.   The pressure distribution shown by this figure is similar to the pressure distribution determined for a beam with uniform pressure in a transverse direction shown in Fig. 5.   However, owing to the effect of deflection of the pad in the circumferential direction, Figs. 9 and 5 are not exactly comparable.   However, these curves are in general agreement and show that peak pressures are reduced with the use of the multiple support.   Increased rigidity of the support structure will result in more uniform distribution of load in a circumferential direction, and further reductions in peak pressures than shown in Fig. 9.   Fig. 10 shows a comparison of the over-all pressure distribution on a sector-shaped pad when resting on either the disk-type or multiple supports.   This figure is effective in showing the advantage of multiple supports in causing a more uniform distribution of load during the starting period.

STRESS PROFILE ALONG CENTERLINE   PROFILE- SECTION AA

STRESS CONTOUR OF A THRUST BEARING SHOE
LARGE DISK SUPPORT

RATED LOAD – 217,000 LBS.

PEAK STRESS – 1430 PSI.

Fig. 8 Static pressure distribution on a sector-shaped pad of a 105-in-diam thrust bearing when resting on a disk-type support. (Ratio of pad radial to circumferential length is 1.64. Disk diameter equals circumferential length of pad at center line of pivot.)

STRESS PROFILE ALONG CENTERLINE   PROFILE, SECTION AA

STRESS CONTOUR OF A THRUST BEARING SHOE
BEAM TYPE SUPPORT

RATED LOAD – 217,000 LBS
PEAK STRESS – 950 PSI.

Fig. 9 Static pressure distribution on a sector-shaped pad of a 105-in-diam thrust bearing when resting on a multiple support. (Ratio of pad radial to circumferential lengths is 1.64. Support locations S/L; see Fig. 5, for optimum static-pressure distribution.)

Pad Distortion Due to Hydrodynamic Pressure of Oil Film

As discussed in the first part of this paper, the desired static loading on a pad is uniform. Furthermore, the pad must be relatively flexible in order for it to conform to small variations in the plane of the runner. After the oil film has been established, and particularly at normal speed, the hydrodynamic pressure on the pad is of parabolic shape with a maximum near the center of the pad. As a result, the arrangement of supporting the pad which is the most desirable from the standpoint of uniform loading during the starting period, is not necessarily the optimum arrangement when operating on an oil film. It is obvious that thrust-bearing-pad distortion, resulting from oil-pressure loading, must

be limited to values smaller than the oil-film thickness in order to avoid metal-to-metal contact and permit the bearing to perform satisfactorily.

The optimum location of supports $(S/L = 0.25)$ for starting conditions, as shown in Fig. 5, corresponds very nearly to a uniform static loading. However, with this location of supports and a change from uniform to parabolic hydrodynamic loading, as shown in Fig. 11, there would result a relatively large deflection of the pad in a concave manner. The magnitude of this deflection can be minimized by either increasing the thickness of the pad or shifting the reactions for a given pad and loading. Fig. 11 shows the deflection of a pad as a function of the support

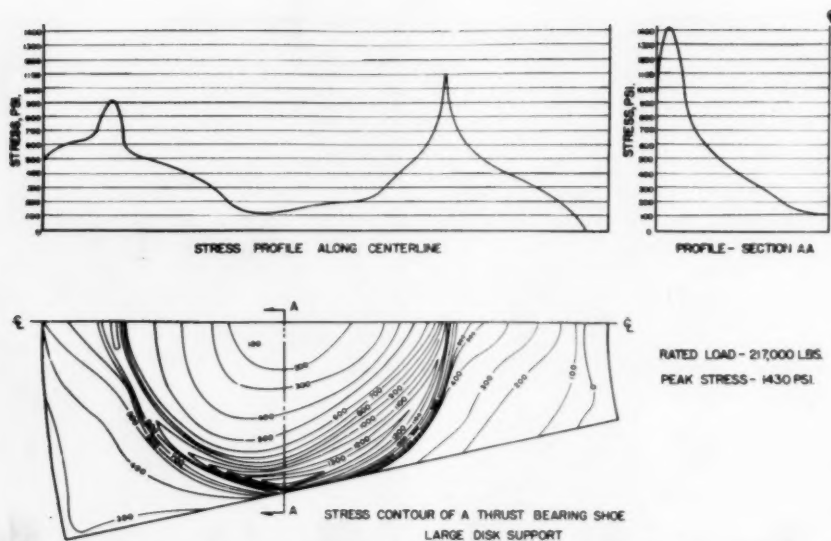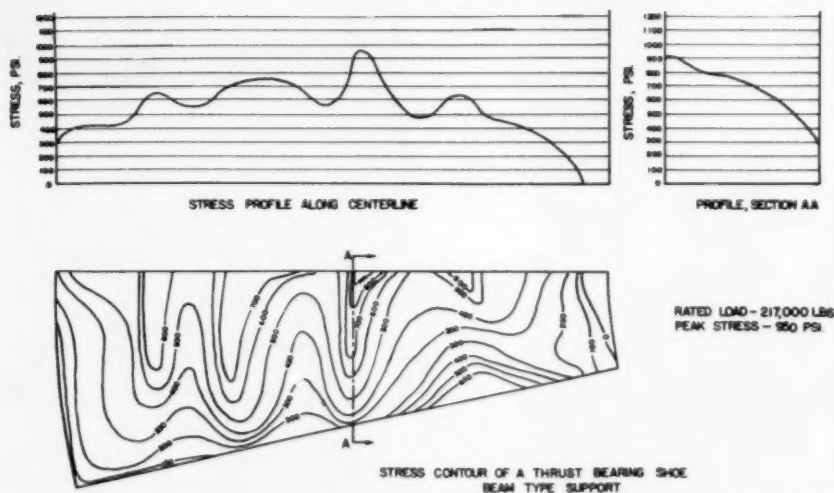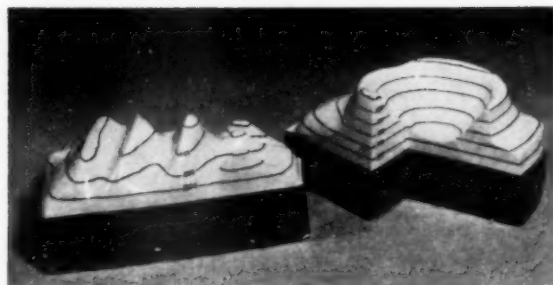Fig. 10 Comparison of static pressure distribution on a sector-shaped pad of a 105-in-diam thrust bearing when resting on a disk-type and multiple support

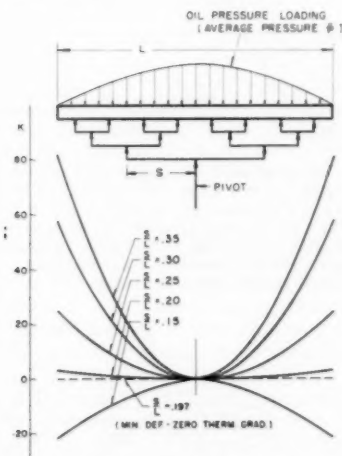Fig. 12 Proportions of pads used in large thrust bearings of vertical waterwheel generators

Fig. 11 Influence of arrangement of multiple supports on distortion along radial center line of a thrust-bearing pad due to parabolic oil-pressure loading. (Based on calculation of a pad of uniform width b where E is modulus of elasticity of pad material, I is moment of inertia of pad cross section. Deflection from a tangent at center line of pivot is $10^{-4} \, KpbL^4/EI$.)

locations. As it is noted, a shift of the (S/L) ratio from 0.25, the optimum for uniform loading, to about 0.20 will minimize the deflection for a given pad and loading when using the multiple support. The latter arrangement is preferable to the use of a thicker pad, since thinner pads conform better to the surface of the runner, as mentioned earlier. The multiple support is particularly well adapted to pads which have large ratios of radial to circumferential lengths, such as shown in Figs. 12(a) and (b). The deflection along the radial center line of such a pad, when resting on the multiple support, shown in Fig. 4, was calculated for a sector-shaped pad of a 105-in-diam, 12-pad thrust bearing. This deflection curve is shown in Fig. 13(a). This pad when resting on the disk-type support, shown in Fig. 2, has the deflection curve shown in Fig. 13(b). As may be noticed from these figures, there is less distortion of the pad when used with the multiple support.

Additional comparison of Figs. 3, 5, 11, and 13 shows that a better compromise between pressure distribution at start, and minimizing distortion during normal operation, is realized with the multiple support. However, it should be noted that Fig. 13 gives relative deflection for pads of the same thickness. The deflection of the pad with the disk-type support can be made as small as that with the multiple support by increasing the pad thickness at the sacrifice of conformability to the runner. On
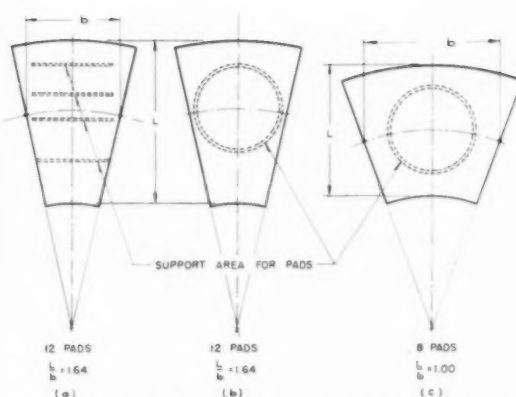
Fig. 13 Deflection along radial center line of a sector-shaped pad of a 105-in-diam thrust bearing due to parabolic oil-pressure loading. (Ratio of pad radial to circumferential lengths is 1.64. a. Resting on a multiple support with optimum ratio S/L. b. Resting on a disk-type support whose diameter is equal to circumferential length of pad at center line of pivot.)

thrust-bearing pads in which the ratio of radial to circumferential lengths are near unity, as shown in Fig. 12(c), the diameter of the disk support can be chosen such as to minimize deflection due to oil-pressure loading and, as will be shown later, to result in small deflections when thermal gradients in the pad are considered. For this reason the disk-type support is used with excellent results on many machines.

Pad Distortion Resulting From Thermal Gradients

In the foregoing discussion of multiple-support location for the pad, no consideration was given to pad thermal gradients. Under normal operation, there can be an appreciable thermal gradient in the pad with corresponding distortion of the surface of the thrust-bearing pad. This thermal deflection can be of the same magnitude as the deflection caused by the oil pressure loading and, therefore, must be considered in the design of large thrust bearings (11, 12).
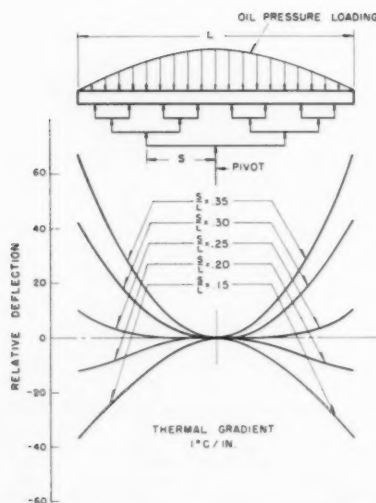
Fig. 14(a) Influence of multiple-support arrangement on distortion along radial center line of a thrust-bearing pad due to parabolic oil-pressure loading and a thermal gradient of 1 deg C per in. pad thickness. (Based on calculations of a pad of uniform width b. For $L^3 = 1000$ and $\bar{p}bL^4/EI = 1$, the deflection is "Relative Deflection" $\times 10^{-4}$. E is the modulus of elasticity of pad material, I is moment of inertia of pad cross section, and $\bar{p}$ is average bearing pressure.)
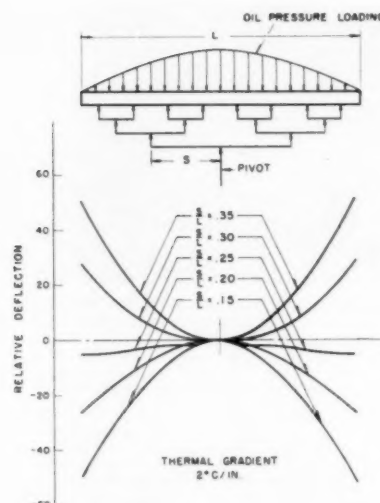
Fig. 14(b) Influence of multiple-support arrangement on distortion along radial center line of a thrust-bearing pad due to parabolic oil-pressure loading and a thermal gradient of 2-deg-C-per-in. pad thickness. (Based on calculations of a pad of uniform width b. For $L^3 = 1000$ and $\bar{p}bL^4/EI = 1$, the deflection is "Relative Deflection" $\times 10^{-4}$. E is modulus of elasticity of pad material, I is moment of inertia of pad cross section, and $\bar{p}$ is average bearing pressure.)
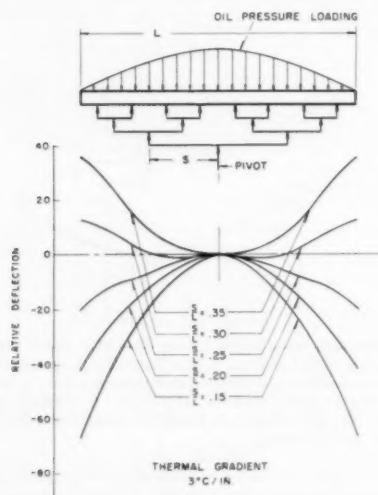
Fig. 15(a) Influence of multiple-support arrangement on distortion along radial center line of a thrust-bearing pad due to parabolic oil-pressure loading and a thermal gradient of 3-deg-C-per-in. pad thickness. (Based on calculations of a pad of uniform width b. For $L^3 = 1000$ and $\bar{p}bL^4/EI = 1$, the deflection is "Relative Deflection" $\times 10^{-4}$. E is modulus of elasticity of pad material, I is moment of inertia of pad cross section, and $\bar{p}$ is average bearing pressure.)
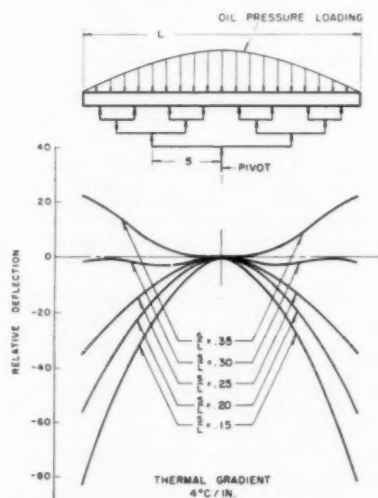
Fig. 15(b) Influence of multiple-support arrangement on distortion along radial center line of a thrust-bearing pad due to parabolic oil-pressure loading and a thermal gradient of 4-deg-C-per-in. pad thickness. (Based on calculations of a pad of uniform width b. For $L^3 = 1000$ and $\bar{p}bL^3/EI = 1$, the deflection is "Relative Deflection" $\times 10^{-4}$. E is modulus of elasticity of pad material, I is moment of inertia of pad cross section, and $\bar{p}$ is average bearing pressure.)

Thermal gradient is a function of shoe geometry as well as of the turbulence, viscosity, temperature, and so on, of the oil in which the thrust-bearing pad is immersed. Thermal gradients of approximately 2 deg C/in. to 4 deg C/in. can be realized depending on the thrust-bearing speed and design.

Figs. 14(a), 14(b), 15(a), and 15(b) show the combined deflection of a thrust-bearing pad due to parabolic oil-pressure loading and various thermal gradients. The optimum ratio (S/L)

can be chosen from these figures to cause minimum distortion of the pad in normal operation. With supports in the optimum location, from the standpoint of oil-pressure loading, as shown in Fig. 11 for the multiple support, there is appreciable deflection when the pad experiences a thermal gradient of 2 deg C/in., as shown in Fig. 14(b). With larger values of thermal gradient, shown in Figs. 15(a) and (b), excessive deflections can exist which
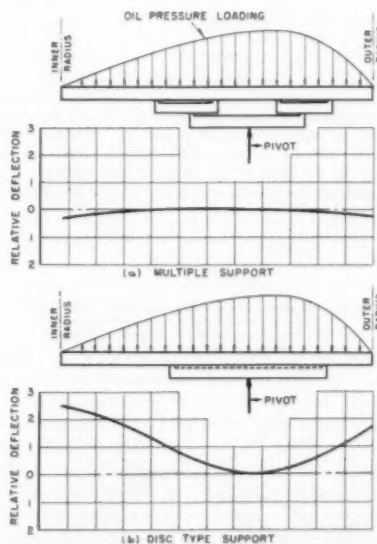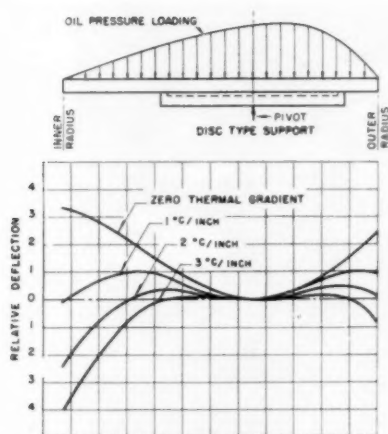
Fig. 16 Distortion along radial center line of a sector-shaped pad of a 105-in-diam thrust bearing due to parabolic oil-pressure loading and various thermal gradients. (Ratio of pad radial to circumferential lengths is 1.64. Pad resting on a disk-type support whose diameter is equal to circumferential length of pad at center line of pivot.)

Fig. 17 Comparison of deflection along radial center line of a sector-shaped pad of a 105-in-diam thrust bearing due to parabolic oil-pressure loading and 2-deg-C/in. thermal gradient. (Ratio of pad radial to circumferential lengths is 1.64. Comparison of cases of pad of same geometry and loading when supported by a disk-type support whose diameter is equal to pad circumferential length at center line of pivot, and when supported by a multiple support with optimum ratio S/L.)

may cause unsatisfactory oil-pressure distribution and possibly solid contact between the pads and runner. Adjustment of support locations to cause minimum distortion of the thrust-bearing pad due to both oil-pressure loading and thermal gradient is necessary in the design of large thrust bearings.

The optimum position of supports may be obtained from curves similar to Figs. 14(a, b) and 15(a, b); then, the pad thickness is chosen to have small peak pressures during starting and minimum distortion due to thermal-gradient effect during normal operation. The requirements of support locations during starting, normal operation, and stopping are such that no support arrangement is optimum for all conditions. It is necessary to choose a support arrangement which is the best compromise.

The effect of thermal gradients of various magnitudes on the distortion of a thrust-bearing pad resting on a disk-type support is shown in Fig. 16. The relative deflections due to oil-pressure loading and thermal gradient are shown by this figure for a sector-shaped pad, of a large 12-pad thrust bearing, having a ratio of radial to circumferential length of 1.64. The maximum diameter of the supporting disk is limited by the circumferential length of the pad near the center of the support. With such a pad, shown in Fig. 12(b), the relative deflection is small for thermal gradients between 1 and 2 deg C per in. For the larger thermal gradient of 3 deg C per in., for example, deflection becomes excessive and the effective surface of the pad is reduced. These thrust-bearing pads with a large ratio of radial to circumferential length can be designed to have small deflections when provided with multiple supports, as shown in Fig. 17. Thrust bearings having a smaller number of pads, hence smaller ratio of radial to circumferential length, shown in Fig. 12(c), can be used effectively with a disk-type support and must be designed for the desired operating condition.

(a) Measurement of Thermal Distortion in Thrust-Bearing Pads. A section of a 105-in-diam thrust-bearing pad 6 in. wide and 6 in. thick, without babbitt, was used for the experimental determination of distortion due to thermal gradients. This section of pad was equal in thickness to the pad of a typical water-wheel-generator thrust bearing. To duplicate the rigidity of the runner and thrust block, the test pad was placed on top of a steel slab 19 in. thick. The load, equivalent to an average static pressure of about 500 psi, was applied to the test section by a

hydraulic ram and clamping rods, as shown in Fig. 18. This load was applied to the test pad through an adjustable multiple support. The pad was provided with water passages drilled near and along the top and bottom surfaces. The desired thermal gradient in the pad thickness was obtained by circulating water at a controlled temperature. Thermocouples, forming a grid of 24 temperature positions in a plane normal to the bearing surface, were inserted into holes drilled 1 in. deep from a side of the pad. The thermocouples were wedged into the holes to assure thermal contact, and the exposed sides of the test pad were thermally insulated to prevent side heat flow and approximate a one-dimensional heat-flow system.

The deflection, or separation between the runner and pad surfaces at the inside radial end of the thrust bearing, was measured with a Tuckerman optical gage. The curvature of the bearing surface of the runner, as influenced by temperature, was determined by six $^{1}/_{10,000}$-in. dial indicators placed at intervals along this surface, shown also in Fig. 18. The total load was measured by electrical resistance strain gages attached to the clamping rods.

(b) Results of Thermal-Deflection Test. Fig. 19 shows the measured deflection of the test pad as a function of the temperature gradients when subjected to static pressure loading as mentioned. The measured deflection is slightly larger than the calculated value as seen from this figure. The difference was found to be caused by a small curvature in the runner in which there is also a small temperature gradient.

(c) Field Observations of Thermal Deflection of a Sector-Shaped Thrust-Bearing Pad. The oil-film thickness of a 105-in-diam thrust bearing was measured in service by means of magnetic strain gages attached to the thrust-bearing pad as shown in Fig. 20. These instruments measured the separation between the pad surface and runner surface at three corners of the pad while the machine was started and stopped during normal operation.

A record of the oil-film thickness, shown in Fig. 21, indicates there is a very rapid increase in oil-film thickness as normal speed is attained. After a few transient changes, the oil-film thickness becomes stable. After the machine was stopped, the oil-film thickness was found to decrease at a very low rate from 0 002 in.
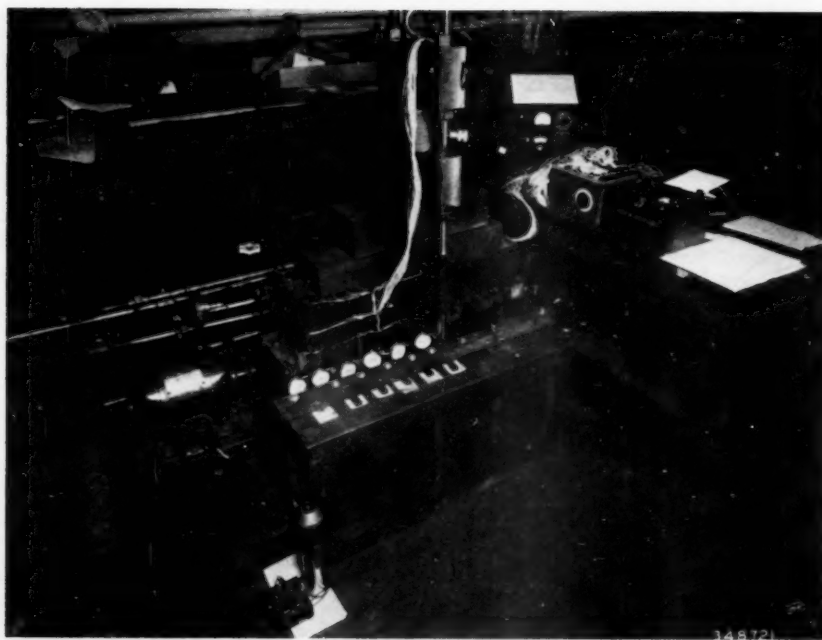
Fig. 18 Test arrangement for determining deflection of thrust bearing pad, under static loading through a variable multiple support, caused by pad thermal gradient. (Deflections observed at inner radial end of a section of a pad of a 105-in-diam thrust bearing.)
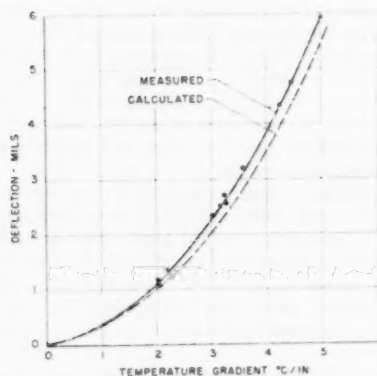
Fig. 19 Deflection at inner radial end of a pad of a 105-in-diam thrust bearing due to pad thermal gradients. (Test section of Fig. 18 loaded 500 psi average static pressure with a multiple support with ratio S/L, set for minimum deflection with parabolic oil-pressure distribution.)

Fig. 20 Sector-shaped pad of a large thrust bearing, ratio of pad radial to circumferential length is 1.00, showing attached magnetic strain gages for measuring oil-film thickness in service

both at the leading and trailing edges of the pad. This remaining separation, or apparent oil-film thickness is much larger than can be accounted for by the normal viscous flow of the oil being squeezed out from between the two surfaces. The 0.002-in. measurement was therefore attributed to remaining separation between the pad and runner caused by the thermal distortion of the pad. In order to obtain a more accurate measure of thermal distortion of the pad, other gages located within the pad are necessary (11). The 0.002-in. deflection measured at the corners of the pad is contributed to by the deflection of the pad in the circumferential direction.

Fig. 20 is a view of the pad used for obtaining oil-film thickness measurement, which was provided with a disk-type support.

There is, therefore, limited correlation between these field observations and the data suggested by Fig. 19, which was for a multiple-supported pad of different ratio of radial to circumferential lengths. However, there is indication, in both these cases, that thermal gradients of sufficient magnitude exist in the pad, making it necessary to consider both load and thermal distortion of the pad in the design of large thrust bearings. As may be noticed in Fig. 20, the marks on the surface of the babbitt indicate that the pad was uniformly loaded during the starting period. The relatively thick oil film, Fig. 21, assured that the distortion of the pad due to load and thermal gradient was not sufficient to result in metal-to-metal contact during normal operation.
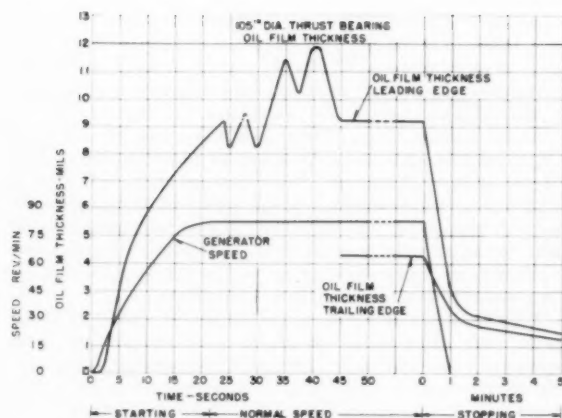
Fig. 21   Oil-film thickness record of a large thrust bearing in service, ratio of pad radial to circumferential length is 1.00, and pad resting on a disk-type support
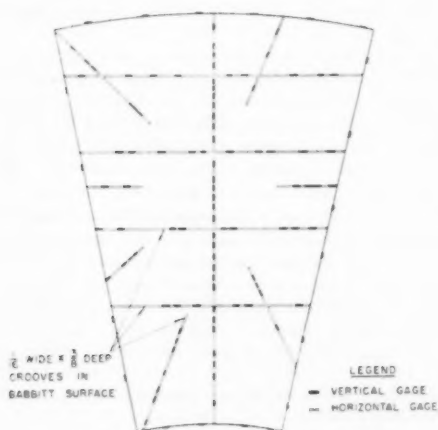
Fig. 22   Diagrammatic location of strain gages used to determine static pressure distribution on the sector-shaped thrust-bearing pad, shown in Fig. 7

In this paper, the support of the pads has been discussed from the standpoint of uniform pressure and no thermal gradients during the starting period. This assumption is justifiable for a bearing which has been at rest for a prolonged period of time; however, when a machine is restarted very shortly after a stop, the effect of thermal gradients should be considered. The pressure distribution resulting from such a condition can be determined with the aid of the beam on elastic-foundation theory. It is found also in this case that the use of the multiple support gives the better compromise for the many conditions of loading and temperature distortions experienced by a thrust-bearing pad.

Conclusions

The performance of thrust bearings can be predicted with reasonable accuracy by means of the known hydrodynamic theory of lubrication. The application of this theory is based on ideal plane surfaces of the runner and pads. In order to approach these conditions, particularly in large thrust bearings, the distortion of the bearing parts due to load and thermal gradient must be considered and kept to a very low value.

Laboratory and field tests have resulted in a better under-

standing of the problem of distortion of the pads and have made it possible to design large thrust bearings of improved reliability and increased capacity.

Bibliography

1 "Performance of Water-Wheel Thrust Bearings During the Starting Period," by C. M. Laffoon, R. A. Baudry, and P. R. Heller, Trans. ASME, vol. 69, 1947, pp. 372–376.
2 "The Theory of Film Lubrication," by R. D. Boswall, Longman, Green & Company, London and New York, 1928.
3 "On Problems in the Theory of Fluid-Film Lubrication With an Experimental Method of Solution," by A. Kingsbury, Trans. ASME, vol. 53, 1931, APM-53-5, pp. 59–75.
4 "Applying Bearing Theory to the Analysis and Design of Pad-Type Bearings," by J. Boyd and A. A. Raimondi, Trans. ASME, vol. 77, 1955, pp. 287–309.
5 "Oil Seals to Provide Positive Lubrication on Large or High-Speed Thrust Bearings," by R. A. Baudry, G. E. Peterson, and G. D. Cooper, published in this issue, pp. 819–825.
6 "Beams on Elastic Foundation, Theory With Application in the Fields of Civil and Mechanical Engineering," by M. Hetenyi, University of Michigan Press, Ann Arbor, Mich., 1946.
7 "Bending of an Infinite Beam on an Elastic Foundation," by M. A. Biot, Trans. ASME, vol. 59, 1937, pp. A-1–7.
8 "Prevention of Babbitt Blisters in Thrust Bearing Pads," by R. A. Baudry, D. W. Gunther, and B. B. Winer, Trans. ASME, vol. 76, 1954, pp. 225–260.
9 "Some Mechanical Considerations in the Design of Large Thrust Bearings," by R. A. Baudry and G. E. Peterson, presented at the ASLE meeting, Detroit, Mich., April, 1957.
10 "Influence of Surface Profile on the Load Capacity of Thrust Bearings With Centrally Pivoted Pads," by J. Boyd and A. A. Raimondi, Trans. ASME, vol. 77, 1955, pp. 321–330.
11 "Études expérimentales sur le fonctionnement des pivoteries industrielles," by S. X. Casacci and A. Peuchmaur, La Houille Blanche, vol. 6, January-February, 1951, pp. 23–42.
12 "Recent Development of Bearings and Lubrication Systems for Vertical Generators," by S. Gynt, Allmanna Svenska Elektriska AB Journal, Vasteras, Sweden, vol. 20, 1947, pp. 72–87.

Discussion

S. M. Denton.[5] The Bureau of Reclamation has in operation many vertical hydraulic turbine-driven generating units with thrust-bearing loadings ranging up to more than 2,000,000 lb. Our experience has shown that loading on thrust bearings has a definite relation to the successful operation of the bearing in that added loading increases the likelihood of bearing failure. This is true despite the apparent nominal unit loading on the bearing surfaces.

The work and studies conducted by the authors should add substantially to the information available regarding large heavily loaded thrust bearings. Disclosure of the behavior of the bearing structure and materials should permit the designer to obtain thrust bearings manufactured to designs and tolerances with considerably better chance of successful operation.

Previously much of the work of adapting the bearings to conditions has been done in the field after erection of the generating unit and largely by the cut-and-try method, since reasons for bearing failures are not always readily apparent.

The authors have analyzed the two sets of conditions which the bearings must withstand; namely, starting and normal operation after starting. Deficiencies in design or manufacture for either condition are most likely to show up during the initial starting, although the starting condition has been considered the most severe. For this reason, the Bureau of Reclamation has adopted the procedure of equipping large heavily loaded thrust bearings with a high-pressure lubricating system to establish and maintain an oil film prior to starting and until sufficient speed is attained to maintain a full-thickness dynamic oil film.

The authors are to be congratulated for this excellent work in the analysis of thrust-bearing performance which should promote confidence in the use of larger hydraulic turbine-driven generating units.

R. B. Jennings.[6] The authors are to be commended for this extensive investigation into the performance of the pivoted-type segmental thrust bearing. It should be recognized that the investigation is for a particular thrust-bearing design. Other thrust-bearing designs may not be subject to the difficulties investigated.

Additional details regarding the thermal-distortion experiments would be helpful to the reader. In particular, was the imposed gradient such that the hot surface was at the babbitt and the cold surface at the back face of the segment? Was the thermal gradient progressing perpendicular to the face of the babbitt? From the description, it is believed that the gradient was perpendicular to the babbitt surface. If this is true, it would be interesting to know whether, in the opinion of the authors, a thinner segment would experience a greater or lesser thermal distortion. Those results are of particular interest which show that the larger the number of support points, the more nearly optimum the pressure profile. The General Electric spring-type thrust bearing embodies the feature of multiple support points but with a greater degree of elasticity, permitting each of these independent support points to vary in elevation consistent with the pressure profile the stiffness of the support member, and the ability of the segment to deform elastically.

D. D. McGregor.[7] The authors have done an excellent job in isolating and evaluating the minute deflections due to load and temperature gradient which occur in the very stiff thrust-bearing pads used on their machines. The close agreement between model studies, mathematical calculations, and checks on operating machines gives assurance that this work has been carefully done and faithfully reported. Searching for deflections of 2 or 3 mils in bearings which carry over 1000 tons of load is truly like looking for a "needle in a haystack." Apparently the search was not in vain.

The fact that the authors find it necessary to isolate and evaluate these very small deflections points to the greatest operating problem with this type of bearing as applied to very large hydroelectric generators. This problem is the low factor of safety on these bearings during the starting and stopping periods. Therefore, from the operator's standpoint, the major question regarding this paper is, how much will the results of this study increase the margin of safety during starting and stopping? Since the design of these thrust bearings admittedly must be a compromise between conditions favoring starting and conditions favoring full-speed operation, any benefits from this investigation must be proportioned between these two conditions. Also, since the distortions are very small, the benefits derived after dividing them should be rather small.

The second question is this, will future designs based on this study allow a relaxation of the requirement that thrust-bearing oil temperature be lowered to 40 C before a unit is restarted? This factor has been troublesome to the operators. At this temperature the oil is better able to carry the loads through the critical starting period and at this temperature the distortion of the thrust-bearing parts is minimized. But the delay of 4 to 7 hr in restarting a unit is costly in terms of power production,

particularly if it occurs during peak power demand. It is true that these units can be restarted sooner if the unit is jacked up so as to establish an oil film on the bearing surfaces. However, in the present-day plant operating personnel is held to a minimum so that operators are not always available for this extra job.

At the present time in the Pacific Northwest, the Corps of Engineers has running, or on order, 21 units fitted with this particular type of bearing. Some of these units are equipped with high-pressure oil pumps which supply oil to the center of each thrust-bearing pad for establishing or maintaining an oil film during starting or stopping of the units. By using this equipment we are now able to operate the machines without the long delay in restarting a unit after a shutdown.

We do not usually advocate the addition of auxiliaries to the main units where failure of the auxiliary can interfere with operation of the main unit. We have been reluctant to change our specification requiring the use of this system but the advantages gained are so pronounced we intend to continue using it on larger machines.

By using the high-pressure-oil starting system the need for making a compromise between starting and running conditions no longer exists. The bearing can be designed on the basis of maintaining its best oil film while running without regard to the starting requirements.

P. C. Warner.[8] The authors have presented a most absorbing paper, emphasizing as it does one aspect of the more practical side of bearing design. The writer is quite interested in the multiple support which seems an ingenious and simple improvement in the area of design where it is applicable. Has this device been employed as yet in a commercial bearing?

While the writer deals with steam-turbine bearings which do not start under appreciable load, and of course are much smaller than the bearings discussed by the authors, yet his experience is similar to that of the authors. Not infrequently when a thrust bearing is in trouble the first signs of distress occur in the center of the shoe. It is difficult to explain this observation unless shoe and collar distortions are considered. While physical

---

[8] Design Engineer, Westinghouse Electric Corporation, Philadelphia, Pa. Mem. ASME.
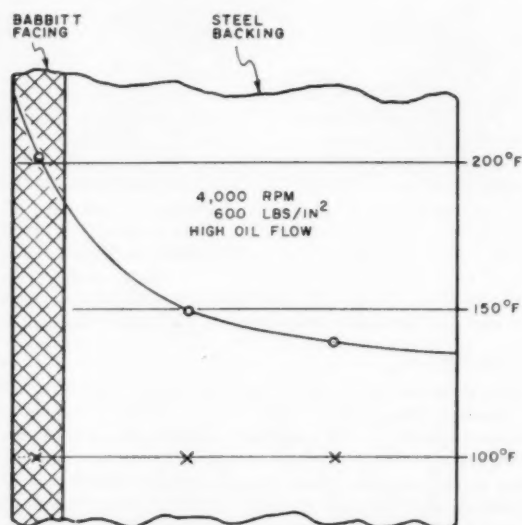
Fig. 23 Temperature gradient near trailing edge of thrust shoe

---

[6] Large Motor and Generator Engineering, General Electric Company, Schenectady, N. Y.

[7] Head, Mechanical Section, Hydro-Electric Design Branch, North Pacific Division, Corps of Engineers, Portland, Ore. Mem. ASME.
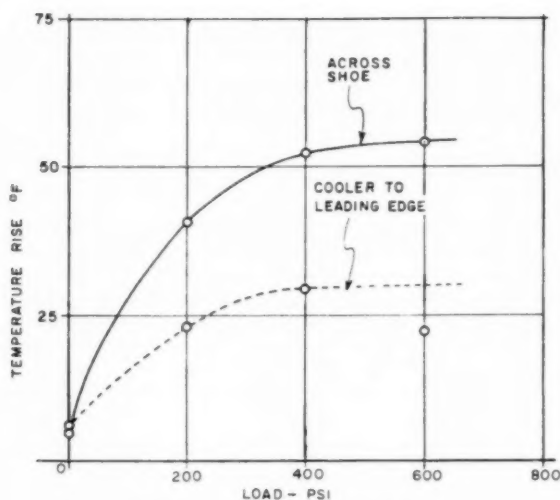
Fig. 24   Temperature rise versus load

sizes are much smaller in turbine bearings, the thermal gradients can be much more severe than those mentioned by the authors. Fig. 23, herewith, gives some typical data in this respect. Gradients of this magnitude can cause quite large distortions even in small bearings. From Fig. 24 it is apparent that, even if the oil going to the shoe is maintained at a low temperature, at the higher thrust loads fairly large temperature gradients can occur.

The writer's experience, although on much smaller bearings in considerably different service, serves to confirm the importance of the matters treated by the authors.

S. E. Weidler.[9]   The authors have presented a very able discussion of thrust-bearing problems, factors, and design considerations.

In comparing various support arrangements the authors used a steel platen to load their test pad. The writer would like some comments on the test method of loading as shown in Fig. 6 of this paper, in comparison to the film loading that occurs in practice. His viewpoint is that, locally, there may be considerable difference in the character of the applied load. Hence, some of the conclusions may be in jeopardy. The writer also would like to caution that the solutions presented in Figs. 11, 13–17, are really first approximations in an iteration procedure. The authors in their setup asked: Given the pressure distribution, what is the distortion? The next question to ask is: Given the film thickness (including distortion), what is the fluid-pressure distribution and so on? The writer's experience in numerical computation leads him to the estimate that at least a 20-step iteration procedure is needed before a solution compatible to the Reynolds, energy, and elasticity equations is determined.

The writer also would like to add the observation that in the bearing field there seem to be two "armed camps"—the experimenter and the analyst. If amicable relationships can ever be set up, the writer is sure that both groups will find it advantageous for several important reasons: (1) The complete thrust bearing is a very complex problem dependent on a myriad of parameters, many of which are interdependent. Thus it is too much to expect an experimenter to sift out one lone effect from a mass of data. (2) Each can provide direction and guidance to the other as to what lines need investigation and

[9] Application Engineer, Analytical Engineering Section, General Electric Company, Schenectady, N. Y.   Assoc. Mem. ASME.

development. For example, experimental techniques in film thickness measurement need refinement. Perhaps cobalt sources would provide a means of determining film thicknesses. (3) Each provides a counterbalance to the other in preventing misapplied reasoning from running amok in the "Never-Never Land" of Peter Pan. Thus, in both groups, there is need of co-operation of one with the other.

Authors' Closure

The authors are indebted to the discussers for their interest in thrust bearing design and their constructive discussions.

Messrs. Denton's and McGregor's experience with the establishment of oil film prior to the starting of the machine is of interest. Many operators feel that the additional factor of safety and convenience of operation are worth the additional equipment.[10] Where such equipment is provided, it is imperative that the oil should be thoroughly filtered; otherwise as shown by recent experience dirt brought between the pad and the runner may cause damage to the bearing surfaces.

For some years it has been the recommendation of the authors' company that the temperature of a thrust bearing be kept as close to 40 C as possible before starting. This recommendation has been made to avoid a repeated number of starts and stops at close intervals and high temperature as could occur during some maintenance operation such as balancing the rotor, adjusting the governor control, etc. This recommendation was made to obtain the maximum factor of safety during operation and should be adjusted to operating requirements. This subject is discussed more completely in another paper.[11] A delay of 4 to 7 hours before restarting a unit as mentioned by Mr. McGregor should not be necessary.

The investigations described in this paper have for object to develop thrust bearings which are able to operate satisfactorily on a wider range of operating conditions.

It is also necessary to design the bearing such that it is possible to start without the use of auxiliary equipment, which might not be available due to failures of mechanical parts or even of the power supply as may be the case on isolated hydroelectric generating units. Mr. Warner's experience with horizontal steam turbine thrust bearings is of particular interest. This type of bearing does not carry any load during the starting period, and yet apparently trouble may be caused by excessive distortion due to temperature gradient. The distortion of a thrust bearing pad caused by thermal gradient or load while operating at normal speed should be appreciably smaller than the oil film if satisfactory operation is desired. This requirement is demonstrated in Figs. 14 and 15 of the paper.

As assumed by Mr. Jennings' discussion, the thermal gradient was perpendicular to the bearing surface. The use of a thin pad with uniform spring loading should be ideal for starting conditions; however, the deflection caused by the oil film loading might become excessive as mentioned in reference (21) of the bibliography which discusses the design of spring-supported thrust bearings. Therefore it is the authors' belief that deflection due to thermal gradient and oil film load are of equal importance on both types of bearings. It is interesting to note that the spring-supported thrust bearing, which formerly was made of a thin disk split in two halves, has recently evolved to a multipad design very similar to the pivoted pad thrust bearing. As mentioned by Mr. Weidler, the relations between the oil-film thickness, oil-film

[10] "Application of Hydrostatic Lubrication to Vertical Waterwheel Generator Thrust Bearings," by J. E. Housley, G. E. Peterson, and T. C. Corey, *Trans. AIEE*, vol. 71, 1952, pp. 206–210.

[11] "Evolution of the Design and Operation of Large Thrust Bearings," by R. A. Baudry, Paper 58-171 presented at the AIEE Winter General Meeting, New York, N. Y., February 7. 1958.

pressure, and distortion of the pad are extremely complicated. The full co-operation of designers and mathematicians with the use of digital computers is necessary for the adequate solution of thrust bearing problems.

The authors' company is making full use of such facilities. In addition, field tests to measure accurately the oil film thickness, the pressure and temperature distribution, and the distortion of the bearing parts are made and found very valuable in obtaining a good understanding and appreciation of the factors involved in the design of large thrust bearings.

Oil Seals to Provide Positive Lubrication on Large or High-Speed Thrust Bearings

By R. A. BAUDRY,[1] G. E. PETERSON,[2] AND G. D. COOPER,[3] EAST PITTSBURGH, PA.

Vertical waterwheel generators are provided with large thrust bearings the surfaces of which are separated by a hydrodynamic oil film. In order to insure the formation of the required oil film, a continuous and sufficient supply of oil is required. With thrust bearings of increased size or speed, aeration of the oil may cause inadequate lubrication of the bearing surfaces or result in oil leakage. This paper discusses the application of positive oil seals which effectively prevent aeration of the oil and oil leakage.

Introduction

WATERWHEEL generators are usually of the vertical type which is the best suited for most hydroelectric power stations. The development of the vertical waterwheel generator to a great extent has been made possible by reliable thrust bearings which support the weight of the rotating parts plus the hydraulic thrust. The surfaces of these bearings are separated at normal speed by a hydrodynamic oil film and thus can operate for many years with practically no wear.

The performance of a thrust bearing can be predicted with reasonable accuracy using the hydrodynamic theory of lubrication which was developed many years ago. However, in order to insure the formation of the required oil film and removal of the friction losses a continuous and sufficient supply of oil is required.

Usually the bearing operates in an oil bath or reservoir, as shown in Fig. 1. The surface of the thrust-bearing runner rests on the thrust pads, the latter being immersed completely in the oil. Rotation of the runner causes the oil to circulate through the cooling coils and between the pads, thus maintaining a continuous supply of cool oil at the inlet of the oil film. This type of lubrication has proved very satisfactory on small or low-speed thrust bearings. However, with the increase in size and speed of bearings, the rotation of the oil in the reservoir induced by the rotation of the runner becomes appreciable and causes the oil surface to assume a parabolic shape depressed at the center, as shown in Fig. 1. If the oil level is thus depressed below the surface of the pads, the oil film no longer can be sustained and wiping of the bearing follows. It is difficult to estimate the oil-level drop at the runner bore for a given bearing because the oil does not rotate uniformly, but Fig. 2 gives one some idea of the parameters involved and the magnitude of the drop that might occur, were the oil free to rotate.

Maintenance of the oil level at a sufficient height above the pads is, therefore, essential. With increase in size and speed of

[1] Manager, Development Engineering, Large Rotating Apparatus Department, Westinghouse Electric Corporation. Mem. ASME.
[2] Mechanical Design Engineer, Mechanical Section, Turbine Generator Engineering Department, Power Apparatus Division, Westinghouse Electric Corporation. Assoc. Mem. ASME.
[3] Engineer, Development Engineering, Large Rotating Apparatus Department, Westinghouse Electric Corporation.

Fig. 1 Low-speed thrust bearing showing variation in oil level caused by rotation of oil in reservoir

$$h = .0256 \cdot N^2 R^2 \text{ INCHES}$$

h = DROP IN OIL LEVEL - INCHES

R = OUTER RADIUS OF ROTATING BODY OF OIL - INCHES

N = REVOLUTIONS PER SECOND

SURFACE SLOPE AT ANY RADIUS -

$$\frac{dy}{dr} = .102 \, N^2 r$$
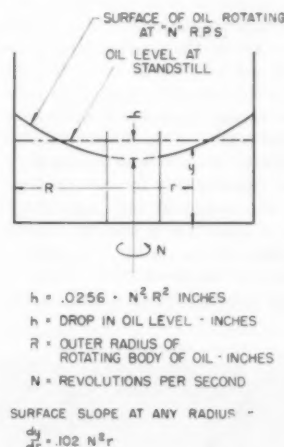
Fig. 2 Radial-pressure variation for uniformly rotating fluid mass. (Using the foregoing equation, it can be shown that for a 40-in-diam thrust bearing with an 80-in-diam reservoir, an oil-level drop, h, of approximately 1.14 in. would take place at the oil guard if the oil were free to rotate uniformly at a speed of 1/8 revolution per sec. In actual modern bearings, the oil is not free to rotate nor is the oil surface free to assume this parabolic shape.)

the bearing the required height of the oil level above the bearing surface becomes excessive. A circular horizontal baffle placed just under the oil level, as shown in Fig. 3, restricts the rotation of the oil to below the baffle and effectively reduces the depression of the oil level at the center of the bearing. The oil pool above the baffle is practically at a standstill and has a stable horizontal surface. The depression of the oil level at the inside diameter of
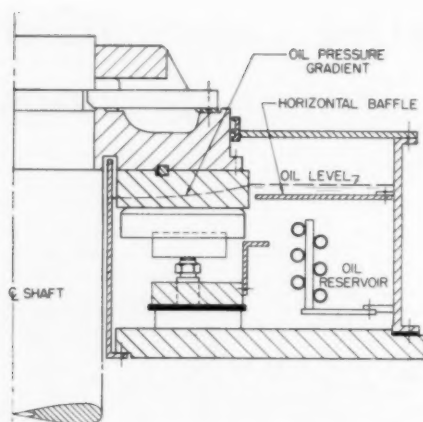
Fig. 3 Medium-speed thrust bearing showing horizontal baffle which controls oil-level gradient

the runner is thus reduced permitting adequate lubrication of larger or higher-speed bearings.

Turbulence and Aeration of Oil at Bearing Center

With a further increase in speed, turbulence begins to appear at the center in the clearance between the oil guard and the bore of the runner. Such turbulence may result first in oil splashing and in more severe cases excessive mixing of air with the oil throughout the reservoir, forming a yellowish emulsion with increased volume. A thick layer of foam is formed on top of the oil level causing oil leakage at the shaft and, as discussed later in this paper the lubrication of the bearing may become inadequate. This entrained air is in addition to the 10 to 13 per cent of air by volume depending on temperature, which is normally in solution in oil and does not change its appearance, viscosity, or lubricating properties. In this condition the air molecules are contained between the oil molecules so there is no change of volume. This dissolved air can be removed only by vacuum treating the oil.

Turbulence at the inner bore may appear at very low speeds unless the inner oil guard has a smooth cylindrical surface concentric with the bore of the runner. Conditions which result in excessive turbulence at low speed are shown in Figs. 4(a) and 4(b). The oil in the gap between the oil guard and the runner rotates
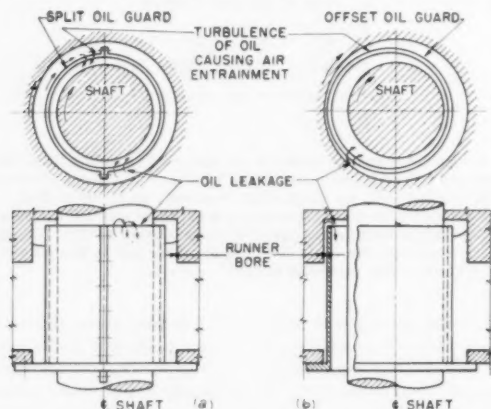
Fig. 4(a) Split inner oil guard which results in turbulence and leakage of oil

Fig. 4(b) Offset cylindrical oil guard which results in turbulence and leakage of oil
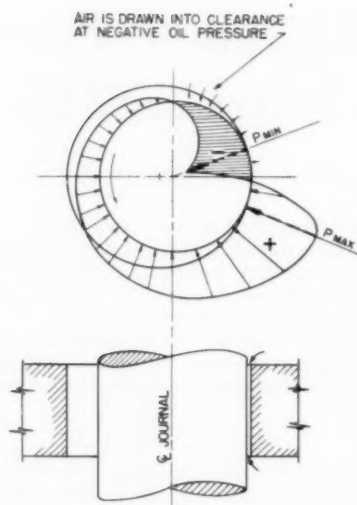
Fig. 5 Pressure distribution in a full journal bearing with insufficient feed pressure. (Taken from "Analysis and Lubrication of Bearings," by Shaw and Macks.)

at approximately one half the peripheral speed of the runner bore. Any obstruction to the motion of the oil will cause it to be thrown above the inner wall, as shown in Fig. 4(a). Therefore, a split inner oil guard should be provided with a smooth cylindrical shield which has a small clearance with the inner bore of the runner. This smooth cylindrical shield also must be concentric with the inner bore of the runner. Otherwise, as is shown in Fig. 4(b), high and low-pressure zones will be produced along the periphery of the oil guard similar to the conditions in a 360-deg journal bearing, Fig. 5. The induced high pressure may be sufficient to cause oil to overflow the oil guard, while in the low-pressure zone air is sucked in and mixed with the oil, thereby aerating the oil throughout the reservoir.

With eccentricity and at relatively low speeds there is little turbulence. With the oil level at sufficient height above the bottom of the runner a smooth oil guard is then satisfactory.

Performance of Inner Seal

In order to determine the conditions which are necessary to prevent oil overflow or aeration at high speed, a thrust-bearing model was built with a transparent oil guard which made it possible to observe the condition of the oil at the inner bore of the bearing under various conditions of speed, oil level, and temperature. This thrust-bearing model is shown in Figs. 6 and 7. The 15-in-diam bearing could be operated at speeds up to 1800 rpm by means of a vertical motor placed above the test rig. The shaft normally located below the thrust-bearing runner was not used in order to permit observation of the oil through the transparent inner oil guard. The condition of the oil at the outer diameter of the runner near the guide bearing also could be observed through several windows placed in the outer wall of the oil reservoir.

It was found that at low speed the circulation of the oil was very stable and there was no aeration. With an increase in speed the oil suddenly became very turbulent, excessive aeration occurring first at the bore and rapidly extending throughout the complete oil reservoir. Large amounts of foam appeared above the oil level, at the inner bore, and at the outer diameter of the runner. At high speed the foam overflowed at the inner oil guard. The magnitude of the overflow increased when speed was reduced or
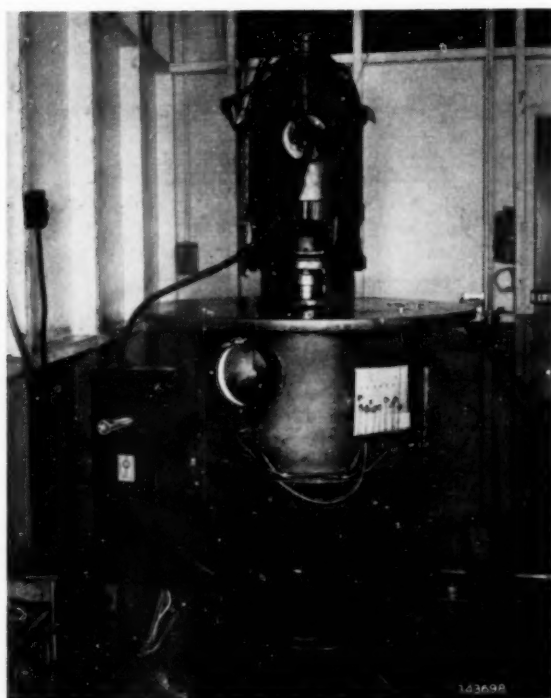
Fig. 6 Thrust-bearing test model used to study oil seals

Fig. 7 Cross section of thrust-bearing test model with transparent inner oil guard to study turbulence

Fig. 8 Cross section of transparent inner oil guard with laminar fluid seal rings

after stopping because the oil aeration at high speed had caused an increase in the oil volume, the oil level at the inner oil guard was no longer depressed by the rotation, and the entrained air bubbles rose to the surface faster, causing a rapid build-up of foam. Similar conditions have been observed on large machines where oil leakage sometimes occurs after an over-speed test.

It appeared that at low speed the oil in the gap between the oil guard and runner bore was moving under laminar-flow conditions. The flow suddenly became turbulent at some speed which was a function of the thickness of the gap and oil temperature. The condition of a viscous fluid between an inner stationary cylinder and an outer rotating cylinder was investigated by Taylor.[4] He established the condition under which the flow becomes turbulent. The conditions in the oil-guard gap are similar to this. Tests made in our laboratory showed that the flow in the oil-guard gap became turbulent when the Reynolds number reached a value between 1000 to 1300.

The Reynolds number R is equal to Vt/ν, where V is the peripheral velocity of the runner bore (in./sec), t is the radial clearance (in.), and ν is the kinematic oil viscosity (in²/sec).

Laminar-Flow Fluid Seal

At high speed the pumping action of the rotating runner is such that the oil level at the inner bore is depressed to near the bearing surface, and turbulence and aeration occur even with the smallest practical radial clearance. Increasing the height of the oil level above the bearing surface tends to raise the speed at which such turbulence can be tolerated. However, design considerations limit this height and other means were found necessary for high-speed bearings.

[4] "Stability of a Viscous Liquid Contained Between Two Rotating Cylinders," by G. I. Taylor, Philosophical Transactions of the Royal Society of London, series A, vol. 223, 1923, 55 pages.
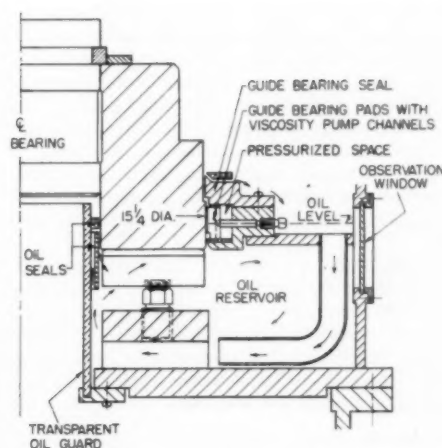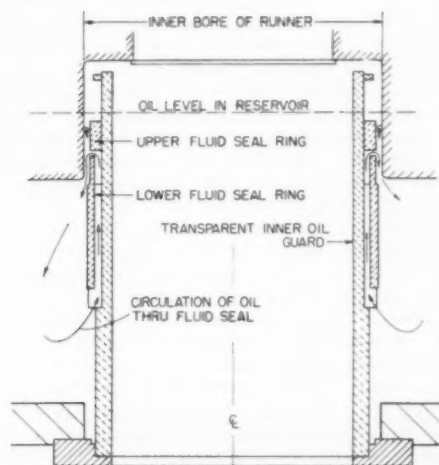
Fig. 9 Transparent inner oil guard with lower seal ring only

A seal ring supported from the oil guard, as shown in Figs. 8 and 9, was found to prevent turbulence and aeration of the oil effectively even at high rotational speed by maintaining a solid film of oil under laminar-flow conditions in the gap between this ring and the bore of the runner. The radial clearance of the seal is chosen so that at the operating conditions the Reynolds number will be below the critical value, as discussed previously.

The space between the seal ring and the oil guard must be sufficient to provide a free flow of oil to the upper level of the seal ring which always must be submerged under oil in order to be effective.

Tests made at high speed on the thrust-bearing model, Fig. 7, showed that the rotation of the runner produced a high-pressure gradient inside the oil reservoir under the horizontal baffle. The oil pool above the horizontal baffle is relatively quiescent and has a stable-free level. It is connected by means of ducts to the part of the main reservoir where turbulence and pressure changes due to rotation are minimum.

Effect of Aeration of Oil

One of the early tests showed that when the thrust-bearing runner was stopped, a large number of air bubbles rose toward the surface of the oil and was trapped under the runner between the pads. Such a large accumulation of air at the inlet to the oil film could be sufficient to interfere with the lubrication of the bearing if it were restarted under that condition. This is demonstrated Fig. 10 where the large volume of air accumulating between thrust-bearing pads immersed in aerated oil can be seen through a glass plate which represents the surface of the runner. The amount of air contained in this oil is representative of conditions sometimes observed in actual thrust bearings. In order to observe the behavior of the trapped air when the thrust-bearing runner is revolving, a glass runner was installed in a 15-in-diam thrust-bearing test rig. A section of this test rig is shown in Fig. 11. It was found that the viscous shear forces resulting from the rotation of the runner were forcing the small bubbles to gather in larger bubbles of rapidly increasing size which adhered strongly to the leading or trailing edges of the pads despite turbulence of the oil which tended to displace them. These bubbles were drawn intermittently into the oil film. Although this test was run under reduced speed and load, it is believed to demonstrate very effectively how entrained air can interfere with the proper formation of the oil
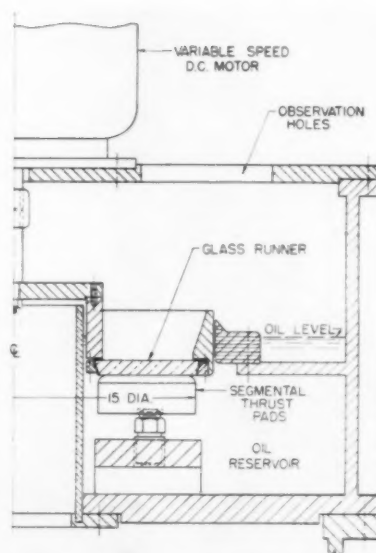
Fig. 11   Thrust-bearing model with glass runner to observe behavior of aerated oil between thrust pads at various speeds

film in a large thrust bearing. It would seem that a thrust bearing starting or operating under such conditions may not receive adequate lubrication.

Although experience has shown that bearings can operate satisfactorily with some aeration, the full extent of the effect of excessive aeration of the oil can only be surmised. The presence of large numbers of small air bubbles in the oil film of a hydrodynamic type of bearing could be highly detrimental to its performance. Compressibility of the air will interfere with the normal formation of the oil film if not prevent it. Furthermore, an air-oil mixture would have a lower conductivity than oil, and would reduce the rate of heat transfer at the surface of the cooling coil and on the bearing parts.

Therefore, prevention of oil aeration is very desirable not only to prevent leakage, but to insure the proper performance of the bearing.
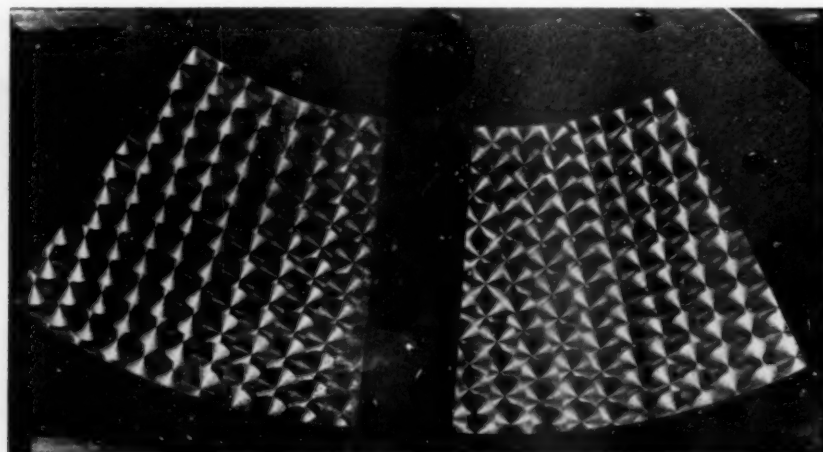
Fig. 10   Illustration showing collection of air bubbles in oil between thrust-bearing pads under a glass plate
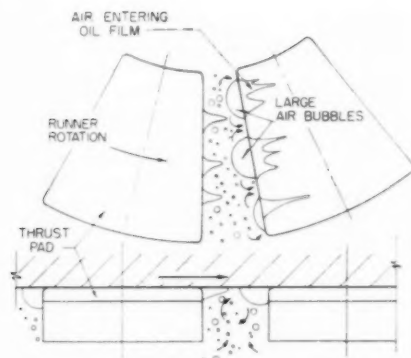
Fig. 12 Air bubbles contained between thrust pads as viewed through rotating glass runner. (Eddies tend to form large bubbles which adhere strongly to leading and trailing edges of thrust pads.)

Fig. 14 Design of viscosity pump in guide-bearing pad

Fig. 13 Combined thrust and guide bearing showing sources of air entrainment in oil

Pressurized Fluid Seals

At the outer periphery of the runner there is much turbulence in the oil, particularly in high-speed machines. The turbulence is further increased when segmental guide bearings are used at the periphery of the runner. This type of guide bearing has been used widely on vertical-type waterwheel generators. It has a very high capacity and permits a very compact arrangement of a combined thrust and guide bearing operating in a common oil bath. This arrangement has been very successful on low and medium-speed machines, and its application has been extended to higher-speed machines. The cross section of a typical combined thrust and guide bearing is shown in Fig. 13. The normal oil level is generally set at about the vertical center line of the guide pads. At high speed the turbulence and splashing which result from this arrangement cause foaming and aeration of the oil throughout the reservoir. This turbulence and the resulting aeration of the oil can be prevented effectively by enclosing the guide bearing in a pressurized housing.

The required pressure can be obtained by the use of viscosity-pump channels machined in the face of the guide-bearing pads, as shown in Fig. 14. This type of viscosity pump has been found to be very simple and effective. The top of the guide-bearing housing is provided with a laminar fluid seal which prevents turbulence and aeration at this location. The clearance of the fluid seal can

Fig. 15 Laminar and turbulent flow of oil contained between thrust runner periphery and guide-bearing seal as a function of clearance, speed, radius, and oil temperature. (These are curves of the equation $h^2 = 42.75 \ (\nu/N^2R)$ which defines laminar to turbulent-flow transition conditions for a viscous fluid contained between two concentric cylinders, with the inner cylinder rotating,[4] where h = radial clearance between cylinders, in., N = revolutions per sec, R = radius of cylinder, in., ν = kinematic viscosity of oil, in^2/sec. These particular curves are for a turbine oil having a viscosity of 200 SSU at 100 F. Three experimental laminar-to-turbulent flow transition points were determined for a seal clearance of 0.020 in. at different values of N^2R. Oil temperatures at beginning of turbulent flow were (A)-65, (B)-61, and (C)-57 C. These points are shown along with dotted portions of the curves for respective oil temperatures.)

be much larger than the clearance at the guide bearing while maintaining laminar-flow conditions and splash-free exit of the oil into the upper part of the oil reservoir.

A number of tests made in the bearing model previously described in this paper, Figs. 6 and 7, permitted the establishment of the requirements necessary for the design of an adequate seal. It was found that the conditions of stability or instability of oil in the seal could be determined accurately by Taylor's criterion for the stability of a viscous fluid contained between an inner rotating cylinder and an outer stationary cylinder.[4] Curves showing the relationship of the parameters involved are given in Fig. 15. Three test points show excellent correlation with the data published by Taylor.

When the seal ring is eccentric with respect to the runner periphery, conditions in the oil film within the seal are the same as for a 360-deg journal bearing with equivalent eccentricity, Fig. 5. In such a narrow bearing the oil film in the low or negative-pressure zone was maintained continuously everywhere around the periphery of the seal. The viscosity pump provided in the guide-bearing pads was found to be entirely adequate to maintain the necessary pressure and flow for the highest speeds expected from this type of combined thrust and guide bearing.

Several large medium-speed combined thrust and guide bearings of this type have been equipped with fluid seals of the type described in this paper, and performance in the field has been entirely as predicted by the laboratory test. These bearings operate without aeration of the oil or any vapor leakage into the generator and are very satisfactory. The cross section of one of these bearings is shown in Fig. 16.
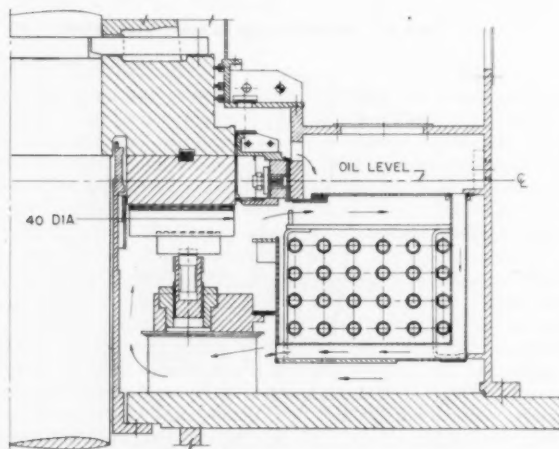
Fig. 16 Medium-size high-speed thrust bearing with an inner oil-guard seal at runner bore and pressurized guide bearing with fluid seal at runner periphery

The same principles can be applied to higher-speed thrust bearings where the high-peripheral speed and resulting high-friction losses prevent the use of a guide bearing at the periphery of the runner. Pressurized seals can be used there as effectively as for lower-speed thrust bearings. As a result of the higher speed and higher operating temperature, however, floating seal rings with smaller clearance are necessary to maintain laminar-flow conditions. Since there is a relatively large radial pressure gradient resulting from the high rotating speed of the runner, it is important that the pressure at the inlet side of the inner seal ring be maintained at a level sufficient to provide the required outward flow and pressure at the whole periphery of the seal even for appreciable eccentricity.

For very high-speed bearings an external positive-displacement pump might prove more desirable than the viscosity pump as used on lower-speed bearings. A bearing of this type is shown in Fig. 17. Here the entire oil reservoir is pressurized by an external pump. Small clearance, floating gland-seal rings are located at the inner and outer diameters of the runner. The flow through these seals returns to the circulating pump. External oil coolers also may be included with this system, which is very similar to a horizontal thrust bearing.

Conclusions

On high-speed machines the oil in the bearing reservoir may be-
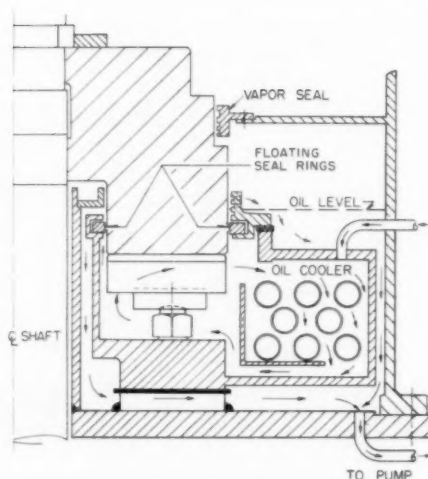
Fig. 17 Very high-speed thrust bearing with floating gland seals at inner and outer diameters of runner and externally pressurized reservoir

come aerated to a considerable extent by turbulence at the bore and the periphery of the thrust bearing. Such aeration of the oil may interfere with the formation of the desired oil film.

The proper application of laminar fluid seals at the bore and periphery of the thrust-bearing runner, and the maintenance of positive pressures at these seals effectively prevent aeration of the oil in large and high-speed thrust bearings, thus assuring adequate lubrication.

References

1 "Recent Developments of Bearings and Lubrication Systems for Vertical Generators," by Sven Gynt, ASEA Journal, vol. 20, April–June, 1947, pp. 72–87.
2 "Foaming and Aeration of Oils in Aviation Power Plant," by R. Tourret, Journal of the Royal Aeronautical Society, vol. 58, January, 1954, pp. 53–60.
3 "Some Results of Tests Made With Segmental Thrust Bearings," by F. Ribary, Brown-Boveri Review, July–August, 1933, pp. 119–122.

Discussion

S. M. Denton.[5] The work represented by this paper adds significantly to the information of the designer for use in designing for successful bearing operation of large and high-speed hydroelectric generating units.

Our experience has demonstrated that the problems involving operation of thrust bearings increase with the loading, size, and speed of the unit. The authors have shown the hazards involved in excessive turbulence and aeration of the oil. Corrective measures have at times been undertaken by the industry in an effort to prevent excessive turbulence in the oil, primarily to avoid overflow of the oil which is sometimes responsible for bearing wiping due to depletion of the oil supply. The authors show that excessive aeration also could be highly detrimental to bearing performance.

The use of oil seals and positive lubrication as suggested by the authors appears to offer correction for excessive turbulence and aeration of the oil so that these need not be limiting factors in the size and speed of thrust bearings. The industry is indebted to the authors for this step toward raising the ceiling on this type of bearing design.

[5] Bureau of Reclamation, Denver, Colo.

F. R. Gould.[6] The authors are to be congratulated for their work in research, development, and testing of oil seals for thrust bearings. It is apparent that the need for this type of oil seal is the result of using a segmental-type guide bearing in conjunction with the thrust-bearing assembly. The change in design from the conventional bearing, which normally is installed between the generator rotor and the thrust bearing, to this type of bearing creates two adverse conditions:

1 The segmental-type guide bearing is operating in the surface oil of the thrust bearing, causing excessive foaming and aeration of the oil.

2 The distance between the thrust bearing and the rotor has been reduced so that the thrust-bearing shaft seal is closer to the rotor and on the suction side of the rotor fans which places it in a low-air-pressure area. Inasmuch as the shaft seal is operating at low pressure, any vapor in the thrust bearing tub can be drawn out past the shaft seal.

All large thrust bearings have many parts to assemble and close tolerances to be maintained in addition to the alignment of the entire rotating unit. The seals described in this paper will have to be installed in the thrust-bearing tub in the upper bracket before the upper bracket is set in place. It is apparent that when the thrust bearing is assembled and the rotor suspended on the bearing, there will be no opportunity to check clearances of the seals. Quite frequently it is necessary to change alignment and shift the upper guide bearing after the rotor has been hung on the thrust bearing. There is no apparent way to shift the seal to compensate for any movement of the shaft.

We had one case of a machine where the oil vapor was being drawn out around the shaft seal due to a low-air-pressure area being created by the rotor fans. This was corrected by installing an air seal around the bearing and connecting it with a duct to the discharge side of the rotor fans so that a positive pressure was maintained at the seals.

It is believed that the design engineers should take another look at the relation of the thrust bearing and upper guide bearing with respect to the rotor fans and the air pressures around the shaft seals. Simplicity in the design of the bearings, oil seals, and baffles will reduce installation costs greatly, and at the same time provide trouble-free operation and the minimum amount of maintenance.

D. D. McGregor.[7] It is interesting to learn the full significance of foaming oil in this type of bearing. The loss of oil through the overflow system and down the shaft, where it can cause endless trouble in collector rings and other generator parts, has always been very evident. However, the ability of these air bubbles to starve the horizontal bearing surfaces seriously has been less apparent. The authors must be congratulated on their thorough investigation of this subject and the imagination used in applying the basic laws of hydrodynamics to the solution

of this problem. It is especially heartening to see positive design action being taken to eliminate foaming so that the brand, type, and grade of oil become less critical. This is important in the Federal service since we have considerably less control over the oil furnished to us than do the private power companies.

It appears questionable to the writer that rotation of the oil in the reservoir is the major cause of the oil depressing at the bore of the runner and rising at the periphery of the reservoir. If this rotation of the oil were the major cause then vertical baffles placed radially would be the answer. However, we find that the job is done by a horizontal baffle, probably for the following reason: This rotating runner acts as a large centrifugal pump with its suction at the runner bore, discharging almost radially at the periphery of the runner. Naturally the suction side will be lower than the discharge by the amount of friction in the system. There is also a velocity head at the point of discharge which is rapidly lost as the oil flows radially in the reservoir, causing a rise of the oil surface in the outer area.

With the horizontal baffle installed the flow is trapped below this baffle so that there cannot be any pile-up of the oil at the periphery of the oil reservoir; therefore, this increases the pressure and oil level of the runner bore.

This paper is very interesting and instructive, particularly to those of us who are concerned with the purchase and operation of these large hydroelectric generators. It again shows that the solution to hydrodynamic problems is best obtained by a combination of model testing and the application of basic theory.

Authors' Closure

The use of vertical baffles placed vertically as suggested by Mr. McGregor would stop the rotation of the oil and prevent the rise of the oil level at the periphery of the reservoir. However, this would also reduce the circulation of the oil through the cooling coil and prevent the effective cooling of the bearing. The friction in the system is of course also contributing to the depression of the oil at the bore.

Experience has shown that in the thrust bearing arrangement shown in Fig. 16 the free rotation of the oil above the cooling coil is effectively used to assist the circulation of the oil through the cooling coil, while turbulence and aeration at the surface of the oil reservoir has been entirely eliminated.

The type of oil seal described in the paper has been installed on several machines. Provisions have been made to facilitate the checking and adjustment of clearances which are kept as large as possible. On high speed machines, in addition to the oil seals, effective air seals, not shown on the figure, are also provided to prevent entrainment of oil vapor in the winding of the machine as discussed by Mr. Gould. This type of air seal is very effective when properly designed.[8]

The authors are indebted to Messrs. Denton, Gould, and McGregor for their interest in the design of thrust bearings and their constructive discussions.

[8] "Modern Bearing Practice for Vertical Waterwheel Generators," by C. M. Laffoon and R. A. Baudry, *Trans. AIEE*, vol. 66, 1947, pp. 363–369.

The Behavior of the Lubricating Film and Side Leakage in Dynamically Loaded Bearings

By M. N. ÖZDAŞ,[1] ISTANBUL, TURKEY

Introduction

THE theoretical treatment of bearings under sinusoidal loads was made by Harrison (1),[2] Swift (2), Dick (3), Burwell (4), and recently by Shawki (5). It was shown by Swift and confirmed by Dick and Burwell that bearings subjected to pure sinusoidal loads are capable of supporting these loads even in the case of no journal rotation. However, in the case of sinusoidal loads with static components, theory shows that the eccentricity will increase continuously and become unity. Consequently the lubricating film will break down. A critical situation occurs when the frequency of the oscillating load is one half the frequency of journal rotation. In this case the theoretical load capacity of the bearing is zero.

One of the assumptions made in the analytical treatments is that a complete film exists around the bearing under dynamic loading conditions. In the present research, it was proposed to investigate the behavior of the lubricating film under static and sinusoidal loads using transparent bearings.

Transparent models have been used often in the past in the study of hydrodynamics. Many techniques have been developed to make the fluid visible using powdered materials in suspension such as graphite or aluminum, dyes of different types, and fluorescent material in the lubricant, with ultraviolet irradiation.

In the field of lubrication, Vogelpohl (6) used graphite powder in a glass bearing 80 mm in diameter and 100 mm wide, in order to observe the breakdown of the oil film in the convergent region. Kolano (7) used a fluid in suspension, such as red-colored kerosene with an addition of glycerine-water solution, which enabled him to take so-called "streak photographs" showing the oil-flow pattern within the bearing. Banks and Mill (8) used medicinal paraffin for the study of cavitation in the nip of two plain transparent cylinders made of plastic material. Dayton and Simmons (9) used a lucite bearing and ultraviolet light for the observation of film breakdown. Red kerosene and a fluorescent oil (Zyglo) mixture were used in an investigation at M.I.T. for the study of cavitation in a 3-in. glass bearing under static load. The most recent work by Cole and Hughes (10) shows the extent of the oil film in a glass bearing 1 in. in diameter, under a static load, using ultraviolet irradiation.

In all the previous work, bearings under static loads were investigated. Preliminary results under dynamic conditions will be presented in this paper.

[1] Doctor, Istanbul Teknik Üniversitesi, Makina Fakültesi.
[2] Numbers in parentheses refer to the Bibliography at the end of the paper.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, August 5, 1957. Paper No. 57—LUB-5.

Nomenclature

The following nomenclature is used in the paper:

d = shaft diameter, in.
b = bearing width, in.
c = diametral clearance, in.
N_j = shaft rotational speed, rps in Sommerfeld number
N_P = load alternations per minute
P = bearing load, lb
p = unit pressure, psi
μ = viscosity, reyns (lb sec/sq in.)
S = Sommerfeld number, dimensionless
$$S = (d/c)^2(\mu N_j/p)$$
Q = side leakage flow, cc/min
p_{oil} = lubricant inlet pressure, psi

Apparatus

The test bearing, Fig. 1, was 2 in. in diameter and 2 in. wide. The test bearing unit (No. 1) consisted of glass bushings with an outer housing made of plastic material which was transparent for observation; the journal of chrome-nickel steel was ground and was supported by two self-aligning double-row ball bearings in a welded housing (No. 2). These ball bearings were axially preloaded to reduce their radial play and consequently provided a journal running accurately with minimum radial displacement under alternating loads. The journal (No. 3) was driven by a variable-speed motor through a V-belt drive. The test bearing was loaded by a spring-steel lever (No. 4) pivoted in the middle and actuated at its other end by an eccentric (No. 5), see Fig. 1(a). The eccentric shaft (No. 6) which was parallel to the journal, was driven independently by a variable-speed motor. Eccentric bushings could be keyed to the other end of the eccentric shaft and a ball bearing was in turn fitted to these eccen-
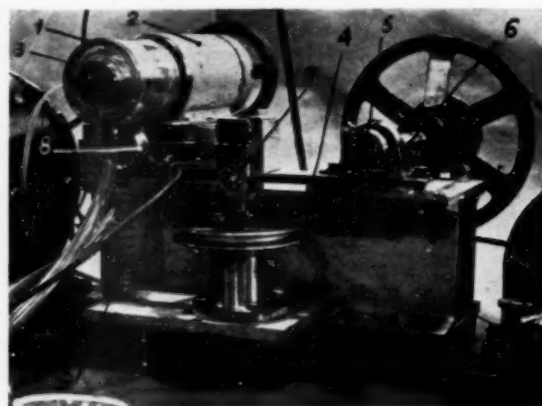
Fig. 1 Test apparatus

tric bushings to reduce the frictional force acting on the spring lever during the loading cycle. The spring lever had a pivot unit at its middle (No. 7). As previously mentioned, one end of the spring lever applied the load to the test bearing through a strain-gage ring dynamometer (No. 8) 2-in. diameter and made of dural. Consequently, the variable load could be measured at its point of application. In order to apply a sinusoidal load with a static component, a helical spring (No. 9) also was fixed to the spring lever at the test bearing end.

During the tests the oil (which contained a fluorescent dye) was supplied either by gravity from a tank or by a gear pump driven independently. The load was recorded continuously as a function of time. A cam-operated switch attached to the eccentric shaft and a solenoid attachment to the camera permitted pictures to be taken at any point of the loading cycle.

Range of Experiments

Tests were carried out with transparent bushings 2 in. in diameter, 2 in. wide, having a diametral clearance of 0.012 in. The speed of the shaft could be varied from 300 to 1500 rpm, that of the eccentric shaft (consequently the load alternation) from 30 to 600 rpm. The amplitude of the sinusoidal load (load line being vertical) also could be varied but the test results presented here are for ±53 lb, see Fig. 2. The oil-inlet pressure could be varied from 0 to 30 psi and the pictures were taken from the positions shown in Fig. 1(a) with an exposure of $1/100$ sec.

The following inlet and grooving arrangements are used:

1 Single oil hole of $1/8$ in. diam.

2 Single oil hole with an axial groove of $1/4 \times 1.45$ in.

3 Single oil hole of $1/8$ in. with two axial grooves of $1/4 \times 1.45$ in. diametrically opposed to each other.
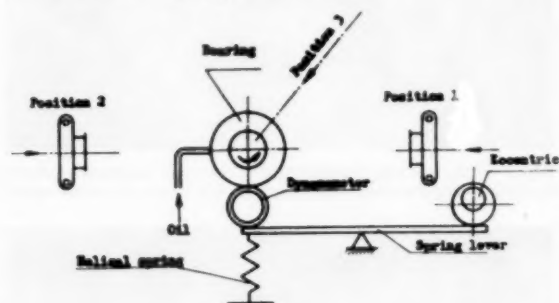
Fig. 1(a)
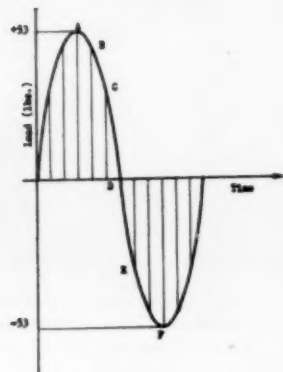
Fig. 2 (a) Loading cycle

4 Single oil hole with a circumferential groove in the middle of the bushing of $0.040 \times 3/8$ in.

The lubricant used in the tests was light mineral oil (with an addition of fluorescent dye) having a viscosity of about 3×10^{-6} reyns (20 cp) at the test temperature.

Film Behavior in Bearings

Under static load conditions when the bearing was lightly loaded (high Sommerfeld number $S > 5$) a complete film was observed. By increasing the load continuously, film breakdown started in the middle of the unloaded region and spread circumferentially. Its extent depended on Sommerfeld number, oil-inlet pressure, and location and type of grooving. The cavitated zone could cover 60 per cent or more of the bearing surface for small S numbers ($S = 0.04$; $b/d = 1$) in bearings having a single-hole oil supply or an axial groove at the horizontal plane, with a vertical load line.

From the complete-film condition, if the load was applied rather quickly, a noticeable evaporation occurred in the cavitated zone after the film breakdown. The film extent and its shape were more or less stable with the exception of the lateral movements of the filaments running through the cavitated zone. The shape of the inlet film depended on the type of oil supply and grooving if the inlet film started from the oil-supply region. (See Fig. 3 for single hole, Fig. 4 for axial groove, Fig. 5 for circumferential groove.) The inlet film had the same shape for both single-hole and axial-groove oil supply if it did not start at the oil-supply point, Fig. 8.

In bushings having a circumferential groove the cavitation occurred as in the previous cases. However, a narrow film existed on both sides of the groove even in the cavitated zone, Fig. 17. The width of this film was quite sensitive to the oil-inlet pressure. In this type of grooving the inlet films were similarly irrespective of oil-inlet location, as the circumferential groove contained lubricant under pressure, feeding the bearing endwise arounds its entire periphery.

On the other hand, the outlet films gave linear or convex breakdown, with thin filaments moving laterally in the cavitated zone in every case.

Figs. 6 and 7 show the cavitated region when the top of the bearing was loaded; while Figs. 12, 14, and 16 show the cavitated region when the bottom of the bearing was loaded, for bushings having a single hole, axial groove, or circumferential groove, respectively.

Under dynamic load conditions (pure sinusoidal) again cavitation occurred, except for light loading. (A continuous film has been assumed by several authors in the analysis of bearings under sinusoidal loads.) Up to the maximum load alternation used in the tests (500 per min), the film could form and reform at the loaded region of the bearing.

Figs. 2(a and b) show the loading cycle under sinusoidal loads. The load line being vertical, let the positive correspond to a load
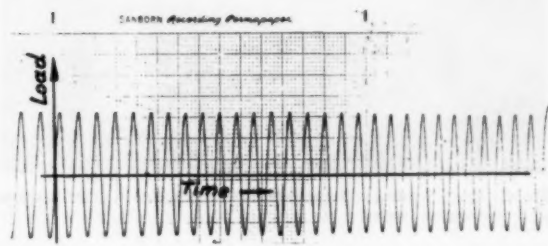
Fig. 2 (b)
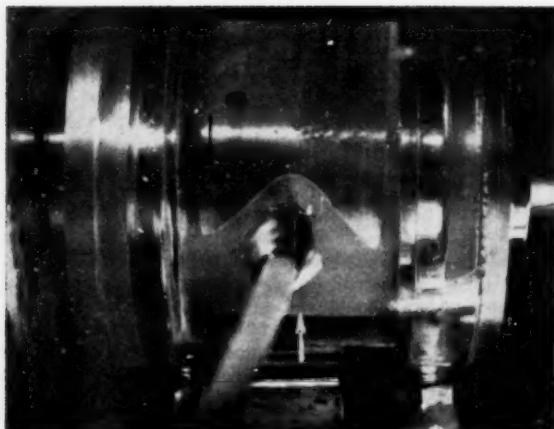
Fig. 3 Static loading; single-hole oil inlet, taken from position 2. (White arrow: Direction of load; black arrow: Direction of journal rotation.)
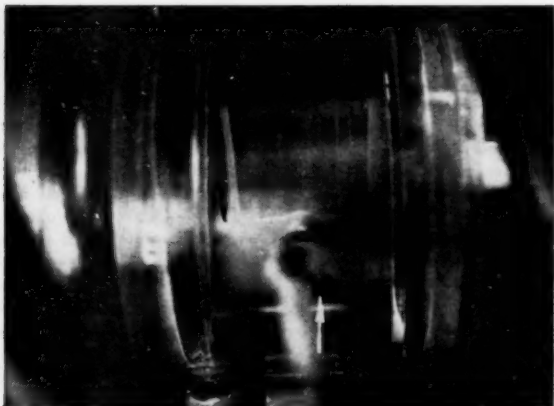
Fig. 4 Static loading; axial groove, taken from position 2
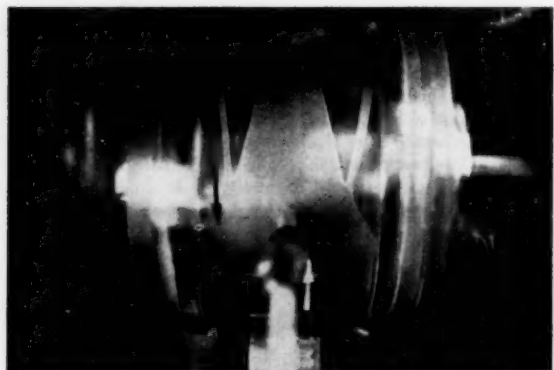
Fig. 5 Static loading; circumferential groove, taken from position 2; lower region loaded
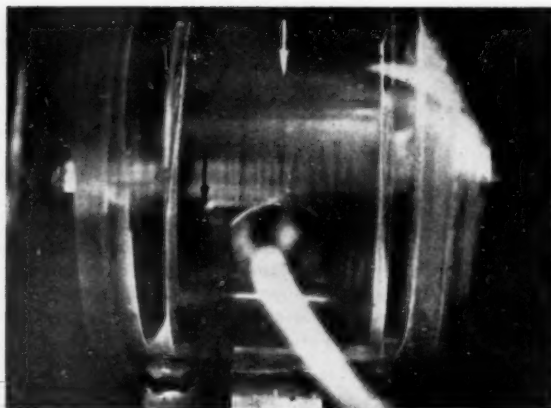
Fig. 6 Static loading; outlet film taken from position 2; for a bushing having an axial groove when top region was loaded
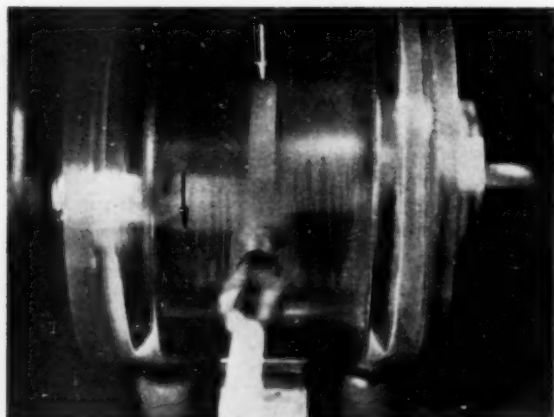
Fig. 7 Static loading; outlet film taken from position 2; for a bushing having a circumferential groove, when top region was loaded
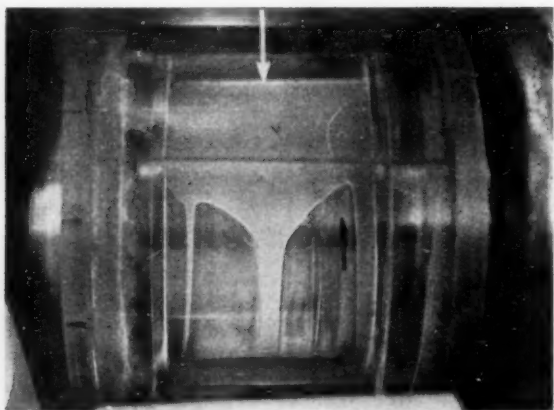
Fig. 8 Dynamic loading; picture taken at point A of loading cycle from position 1; $N_i = 500$ rpm; $N_P = 50$ per min; single-hole oil supply
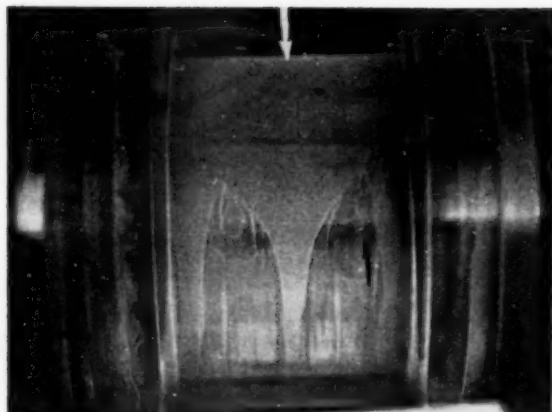
Fig. 9   Dynamic loading; picture taken at point B of loading cycle from position 1; $N_j = 500$ rpm; $N_P = 50$ per min; single-hole oil supply

Fig. 12   Dynamic loading; picture taken at point F from position 1; $N_j = 500$ rpm; $N_P = 50$ per min; single-hole oil supply

Fig. 10   Dynamic loading; picture taken at point C of loading cycle from position 1; $N_j = 500$ rpm; $N_P = 50$ per min; single-hole oil supply

Fig. 13   Dynamic loading; picture taken at point A from position 1; $N_j = 500$ rpm; $N_P = 50$ per min; double axial groove

Fig. 11   Dynamic loading; picture taken at point E from position 1; $N_j = 500$ rpm; $N_P = 50$ per min; single-hole oil supply

Fig. 14   Dynamic loading; picture taken at point F from position 1; $N_j = 500$ rpm; $N_P = 50$ per min; double axial groove

Fig. 15 Dynamic loading; picture taken at point A from position 1;
N_j = 500 rpm; N_P = 50 per min; circumferential groove

Fig. 16 Dynamic loading; picture taken at point F from position 1;
N_j = 500 rpm; N_P = 50 per min; circumferential groove
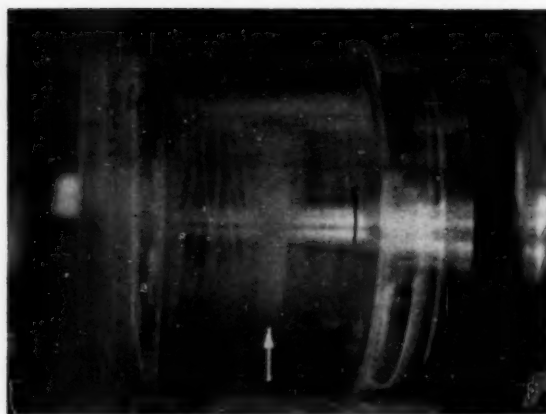
Fig. 17 Static loading; picture taken at point A from position 3;
lower region loaded

directed downward, and negative, to a reversed load directed
upward. In this case the top and the bottom regions of the bear-
ing will be loaded successively during one cycle. Figs. 8, 9,
and 10 show the cavitated and film-covered zones in a bushing
having a single hole (same for axial groove) taken at the points
A, B, and C of the loading cycle from position 1, Fig. 1(a). In
these pictures the top part of the bearing was loaded. It is in-
teresting to note that the thick oil filament which supplied
lubricant to the bearing had to run through the cavitated zone
in order to form an inlet film. Figs. 11 and 12 show the lubricant
film when the loading cycle was at E and F, respectively. In
these cases as the load was reversed the lower regions were loaded
and the linear outlet films are clearly seen.

Figs. 13 and 14 show the behavior of the lubricating film when
the top region (at point A) and the lower region (at point F)
were loaded, respectively, during the loading cycle for a bushing
having a single oil hole with two axial grooves diametrically
opposing each other. It is seen that the additional axial groove
does not disturb greatly the shape of the film.

The cavitated regions in a bushing having a circumferential
groove are shown in Fig. 15 when the top region was loaded
(point A) and in Fig. 16 when the lower region was loaded during
the cycle. Fig. 17 taken from position 3, Fig. 1(a), shows the
cavitated region when the lower region was loaded.

Under dynamic conditions, except very low load alternations
(20–30 per min) a complete film could not be observed at the
zero load, point D of the cycle, as there was not enough time for
the lubricant to cavitate in one region, to form a complete film,
and then to cavitate in the opposite region. Therefore under
these conditions the load-carrying film rotated intermittently
within the clearance space in the direction of shaft rotation
with a speed equal to the alternation of the load. The effect
of this behavior on frictional characteristics of dynamically
loaded bearings should be investigated further. On the other
hand, it was difficult to obtain pictures at the point D because
the load is there changing its direction and, in addition, the
inclination of the sine curve is high at that point.

At alternations higher than 400 per min it was impossible to
obtain proper pictures. No foaming was observed during any
of the tests, even up to the maximum peripheral speed used
(4 m/sec).
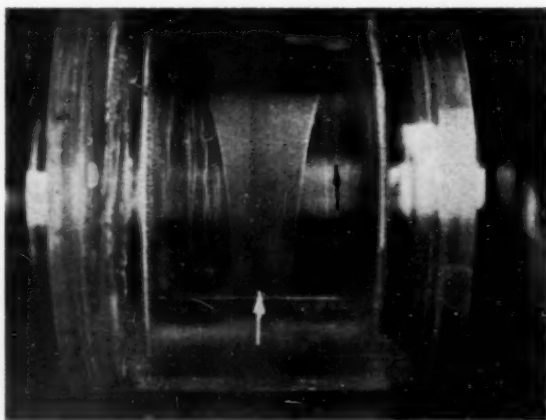
End Leakage Tests

Preliminary tests were carried out to determine the end leak-
age from bearings subjected to sinusoidal loads.

Figs. 18, 19, and 20 show typical end-leakage curves in bearings
having single-hole oil entry, axial groove, and circumferential
groove, for different journal speeds, load alternations, and oil-feed
pressures. It is interesting to note that each leakage curve has
a minimum at a load alternation which corresponds to one half
the journal speed. In Figs. 21 and 22 the same data are plotted
by taking N_P/N_j ratio as abscissa. From these figures it can be
seen that there is a correlation of the data and all curves have
the same behavior. End leakage decreases up to the value
N_P/N_j = 0.5, with a minimum at this critical value, and then
increases again for N_P/N_j > 0.5. This ratio is also critical for
the load capacity of bearings subjected to sinusoidal loads.
When the load is alternating at a frequency equal to one half the
journal speed, the load capacity is theoretically zero, and it has
been shown experimentally by Dayton and Simmons (9) that
in this case the load is carried at a high eccentricity [see also
Shawki (5)]. The end-leakage curves in Figs. 21 and 22 are also
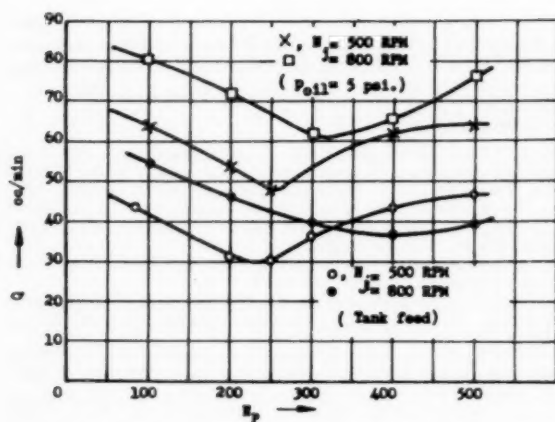
Fig. 18 End-leakage flow rate for different load alternations with single-hole oil supply
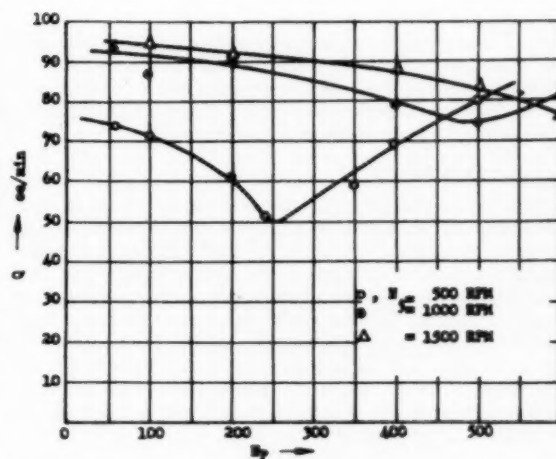
Fig. 19 End-leakage flow rate for different load alternations from an axially grooved bearing
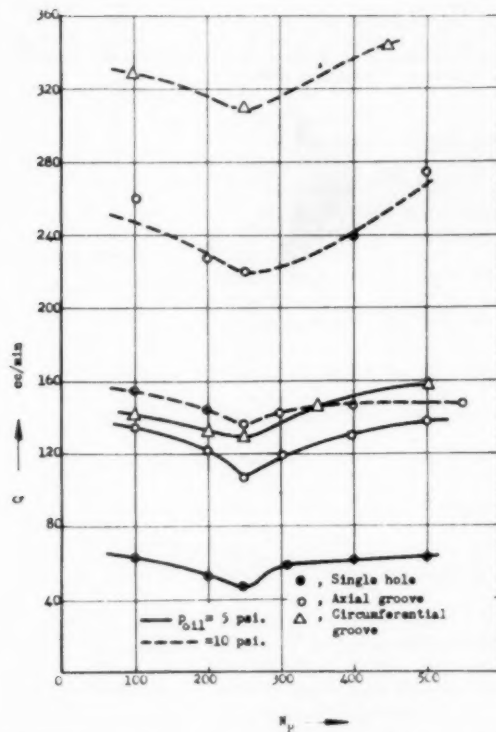
Fig. 20 End-leakage flow rate for bushings having single hole, axial groove, and circumferential groove for different oil-inlet pressures and load alternations ($N_j = 500$ rpm)
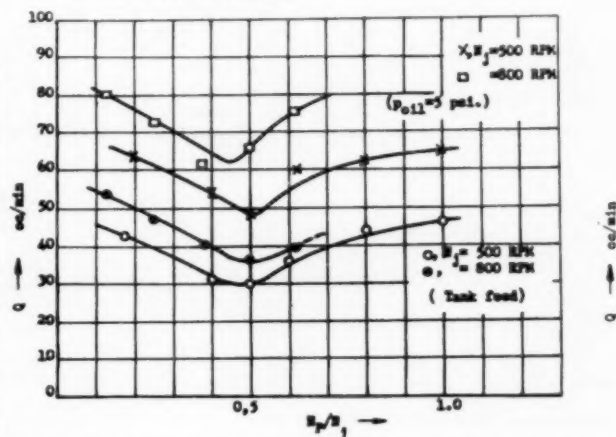
Fig. 21 End-leakage flow rate plotted against N_P/N_j ratio; single-hole oil supply
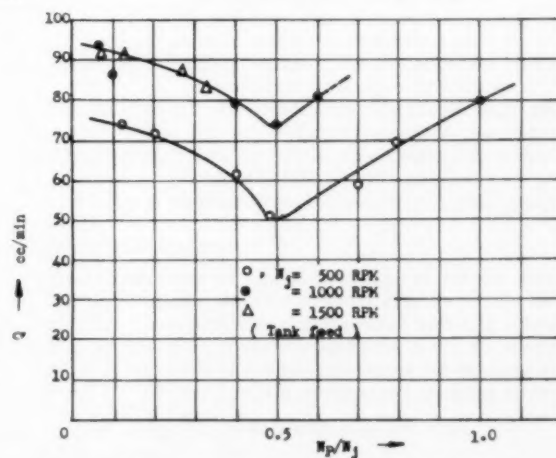
Fig. 22 End-leakage flow rate plotted against N_P/N_j ratio in a bushing having an axial groove

an indication of the critical behavior of a bearing at $N_P/N_j = 0.5$. Theoretically in bearings subjected to pure sinusoidal loads the hydrodynamic pressure is zero for $N_P/N_j = 0.5$. Therefore the end leakage at this critical ratio is due only to oil-inlet pressure. Further investigations of oil leakage may yield useful information for bearings under dynamic loads.

Summary and Results

This experimental investigation was carried out in order to determine the behavior of the lubricating film in bushings having oil entry by means of a single hole, an axial groove, or a circumferential groove, under static or sinusoidally varying load.

Cavitation occurred, except for lightly loaded bearings, under static or dynamic loading conditions. Under sinusoidal loads, up to the maximum load alternation used in the tests (500 per min), the film could form and reform at each loaded region of the bearing. In doing this the load-supporting film moved intermittently within the clearance space in the direction of shaft rotation at a speed equal to the alternation of the cyclic load. The effect of this behavior on bearing frictional characteristics should be investigated.

The side-leakage flow from the bearing was found to depend on the ratio N_P/N_j, and a minimum in side leakage was observed at the value $N_P/N_j = 0.5$, which is also critical from the point of view of load capacity.

Acknowledgments

This research was carried out at Massachusetts Institute of

Bibliography

1 "The Hydrodynamical Theory of Lubrication of a Cylindrical Bearing Under Variable Load and a Pivot Bearing," by W. J. Harrison, *Trans. Cambridge Philosophical Society*, vol. 22, 1919, p. 373.

2 "Fluctuating Loads in Sleeve Bearings," by H. W. Swift, *Journal of the Institution of Civil Engineers (London)*, vol. 5, 1937, pp. 161–195.

3 "Alternating Loads on Sleeve Bearings," by J. Dick, *Philosophical Magazine*, vol. 5, 1944, p. 841.

4 "The Calculated Performance of Dynamically Loaded Sleeve Bearings," by J. T. Burwell, *Journal of Applied Mechanics*, TRANS. ASME, vol. 69, 1947, pp. A-231–245.

5 "Analytical Study of Journal-Bearing Performance Under Variable Loads," by G. S. A. Shawki, TRANS. ASME, vol. 78, 1956, pp. 457–464.

6 "Beitrage zur Kenntnis der Gleitlagerreibung," by G. Vogelpohl, *Zeitschrift, Vereines deutscher Ingenieure*, Forschungsheft 386, 1937.

7 "Study of Lubricant Flow in Bearings, by Streak Photography," by F. J. Kolano, *Product Engineering*, vol. 24, no. 10, 1953, pp. 162–163.

8 "Some Observations on the Behavior of Liquids Between Rotating Rollers," by W. H. Banks and G. C. Mill, *Proceedings of the Royal Society (London)*, vol. 223, 1954, pp. 414–419.

9 "Hydrodynamic Lubrication of Cyclically Loaded Bearings," by R. W. Dayton and E. M. Simmons, NACA TN 2544.

10 "Oil Flow and Film Extent in Complete Journal Bearings," by J. A. Cole and C. J. Hughes, The Institution of Mechanical Engineers, Preprint, February, 1956.

Interpreting Dynamic Measurements of Physical Systems

By SIDNEY LEES,[1] CAMBRIDGE, MASS.

Part I—Aims and Objectives

Background concepts are developed for the interpretation of dynamic measurements. General objectives for a dynamic-measurements program are discussed together with suggested criteria for selecting a test method. A generalized concept of the linear system is developed that has its quantitative form in the idea of the multicoupled operating system with several inputs and outputs. Several schemes to represent the parameters and coefficients of the performance equations associated with the multicoupled operating system are given, according to the test procedure. A general discussion is included of the experimental limitations bounding the interpretation techniques.

Nomenclature

THE following nomenclature is used in the paper:

Quantity Symbols

q_{in} = input quantity

$q_{in(d)}$ = dth input quantity of a multi-input system

q_{out} = output quantity

$q_{out(c)}$ = cth output quantity of a multi-output system

In general, $q_{out(c)} = q_{out(c)(ts)} + q_{out(c)(fs)}$

$q_{out(c)(ts)}$ = transient-solution part of the cth output quantity

$q_{out(c)(fs)}$ = forced-solution part of the cth output quantity

$(Sg)q_{in(d)}$ = signal representing the dth input quantity

$(Sg)q_{out(c)}$ = signal representing the cth output quantity

$q_{in(d)(ind)}$ = indicated value of the dth input quantity

$q_{out(c)(ind)}$ = indicated value of the cth output quantity

t = time

T = period or duration depending on the context

(r) = receiver

$q_{out(c)(meas)}$ = measured output quantity (Shinbrot's method)

$(E)q_{out(c)}$ = error of cth output quantity

$= q_{out(c)} - q_{out(c)(meas)}$

$(ESC) = [(E)q_{out(c)}]^2$ = error square condition

$\Delta q_{out(c)}$ = variation of the cth output quantity

$(D)q_{in}$ = deviation of input quantity

$(D)q_{out}$ = deviation of output quantity

Transient Response Symbols

Δ = characteristic equation determinant associated with a multicoupled operating system

Δ_{cd} = characteristic equation determinant with the cth column replaced by the operator coefficients of the dth input

$A_{i(c)}$ = coefficient associated with the ith characteristic equation root transient solution term of the cth output

$A_{i(c)(1)}$ = special symbol in Shinbrot's least squares method

$= $ 1st approximation to $A_{i(c)}$

$A_{i(c)(2)}$ = 2nd approximation to $A_{i(c)}$

$\Delta A_{i(c)}$ = variation of $A_{i(c)}$

Parameter Symbols

$z_i = x_i + jy_i$ = ith characteristic equation root

x_i = real part of the ith root

y_i = imaginary part of ith root

$j = \sqrt{-1}$

τ_i = characteristic time associated with ith root = $-1/x_i$

$\omega_{n(i)} = \dfrac{2\pi}{T_{n(i)}}$ = angular undamped natural frequency associated with ith characteristic equation complex root

$T_{n(i)}$ = undamped natural period associated with ith characteristic equation complex root

ζ_i = damping ratio associated with ith characteristic equation complex root

$A_{\zeta i} = \cos^{-1}\zeta_i$ = angle of damping ratio associated with ith characteristic equation complex root

$(bp)_{(1st\ ord)}$ = 1st order breakpoint associated with real characteristic equation roots = $-x = 1/\tau_i$

$(bp)_{(2nd\ ord)}$ = 2nd order breakpoint associated with complex characteristic equation roots = ω_n

Δz_i = variation of ith characteristic equation root

Differential Equation Symbols

$a_{i(cs)}$ = coefficient of the ith derivative of the cth input quantity in the eth equation of the equation set associated with a multicoupled operating system

$b_{k(ds)}$ = coefficient of the kth derivative of the dth output quantity in the eth equation of the equation set associated with a multicoupled operating system

$p = \dfrac{d}{dt}$ = derivative operator with respect to time

$p^i = \dfrac{d^i}{dt^i}$ = ith derivative operator with respect to time

Functional Symbols

$[PF]_{(mos)[d; c]}$ = performance function associated with a multicoupled operating system relating the dth input quantity to the cth output quantity

(When no ambiguity exists, the subscripts may be dropped.)

$[PF]_{(mos)[d; c]}(p)$ = performance function written in terms of the variable $p = d/dt$

$[PF]_{(mos)[d; c]}(j\omega)$ = performance function written in terms of the variable $j\omega$

$[WF]_{(mos)[d; c]}(t)$ = weighting function associated with a multicoupled operating system relating the dth input quantity to the cth output quantity. The weighting function integral is a special form of the performance function.

$[FT]F(t)$ = Fourier transform of the functions $F(t)$

$S_{(mos)[d; c]}$ = sensitivity associated with a multicoupled operating system relating the dth input quantity to the cth output quantity

$S_{(he)}$ = sensitivity of heat exchanger

S_{dem} = demodulator sensitivity

$\sum_i$ = summation over index i

$\prod_i$ = product over index i

(CF) = correction function (for Fourier transform approximations)

$[ACF]_{q_{in}}$ = autocorrelation function of the input quantity

$[CCF]_{[q_{in}; q_{out}]}$ = cross-correlation function of the input quantity to the output quantity

$[NST]$ = number series transformation

$[IF]$ = impulse function

$[IFS]$ = impulse function strength = $\int_{-\infty}^{\infty} [IF]dt$

(MSE) = mean square error = $\frac{1}{2}T[(E)q_{out}]^2$

$P_n(j\omega)$ = numerator polynomial in the variable $(j\omega)$

$P_d(j\omega)$ = denominator polynomial in the variable $(j\omega)$

Steady-State Frequency Response Symbols

ω = angular forcing frequency

$q_{in(d)(a)}$ = amplitude of the dth sinusoidal input quantity

ω_ν = νth value of angular forcing frequency

$q_{out(c)a}$ = amplitude of cth output quantity

$\phi_{(d-c)}$ = relative phase of $q_{out(c)}$ with respect to $q_{in(d)}$ for frequency ω_ν

Symbols Associated With Special Examples

Aircraft symbols

m = mass of aircraft

S = wing area

U = mean forward velocity

ρ = air density

c = wing chord

I_y = aircraft moment of inertia about its y-axis

q = $\frac{1}{2}\rho U^2$

w/U = ratio of vertical velocity to horizontal velocity

θ = angle of pitch

δ_e = change in elevator deflection

$C_{z_a}, C_{z_{\delta_e}}, C_{m_a}, C_{m_q}, C_{m_{\dot{a}}}$ are stability derivatives (quantities to be determined)

Electrical symbols

R = resistance

L = inductance

C = capacitance

i = current

e = voltage

Symbols Used in the Figures

Parameter	Symbol in self-defining notation	Working symbol
Undamped angular natural frequency	W_n	ω_n
Damping ratio	DR	ζ
Low-damping ratio characteristic time	$(CT)_{(ldr)}$	τ
Larger high-damping ratio characteristic time	$(CT)_{(hdr)(larger)}$	τ_L
Smaller high-damping ratio characteristic time	$(CT)_{(hdr)(smaller)}$	τ_2
High-damping ratio characteristic time	$[(CT)R]_{(hdr)}$	ν

TP = transient peak

DVR = dependent variable ratio

ext = extended

TPR = transient peak ratio

m, m_r = order integers of transient peaks

T_p = pulse duration

ωT_p = forcing frequency-pulse duration product

μ = nondimensional amplitude ratio

Introduction

The occasion frequently appears when it is necessary to obtain experimental information concerning the performance of a particular physical system. At this point, before any work is done, it is necessary to decide (a) what kind of tests to apply, (b) how the tests are to be carried out, and (c) how to interpret the results. Ideally, these questions should be answered during the planning stage. Frequently, under the pressure of getting information, some tests are performed and the results interpreted by whatever techniques may be available. Then it is discovered that most of the data are grouped in one small part of the operating range, and more tests are needed before significant conclusions may be developed. Sometimes it turns out that the extension of the test procedure is very costly or even impossible to execute in the required region. Either a new test procedure must be devised or the effort to obtain experimental information must be abandoned.

The basis for making a choice of one type of test and technique for interpretation rather than another is probably more a matter of prejudice than of a rational decision. There does not appear to be an existing methodology upon which to base an unbiased choice. Whenever a new method or technique is discovered, it is touted as the answer to all problems by its discoverers, while the practitioners of the older methods continue that which they know best by usage. Some efforts have been made to compare the different procedures that are being used but those making the comparison will surely concede that much more remains to be done in this area.

This is the first of a series of papers on the broad theme of interpreting dynamic measurements. While the author lays no claim for omniscience, it is his modest hope to uncover the existence of the problem and perhaps to lay out its extent. In this first part, it is intended to define the problem and the concepts

associated with it. In a second part, the methods for testing physical systems and interpreting the resultant dynamic measurements are organized into one specific pattern, not necessarily the only one. The evaluation of the methods will be taken up in later papers.

Purposes of Dynamic Measurements

Physical systems are tested for a wide variety of reasons that may be grouped in a general way under the following headings: (a) To acquire information as to the performance of an existing system. (b) To determine if a given system is performing as predicted. (c) To acquire information for improving an existing system. (d) To search for defects in system performance. (e) To acquire information for designing a new system. (f) To determine the system parameters.

By a system is meant a physical entity, complete enough to be identified as a unit. It may be part of a larger entity, in which instance it is described as a subsystem or a component. Size or complexity is irrelevant to the discussion. The list of reasons is neither exhaustive nor mutually exclusive. The most elementary tests may be applied to measure performance quality in order to determine, for example, the order of magnitude of sensitivity, or uncertainty, or range, or speed of response, or the influence of interfering quantities. When the tests are refined, it may be found possible to evaluate the data more exactly. The improvements may give greater confidence to quantitative relationships that have been established among the system variables and parameters, or, as is often the case, the additional refinement may require a re-examination to search out neglected influences. A test program that purports only to measure the quality of performance may have to determine the system parameters in order to express the performance quality most effectively.

It is generally necessary to carry out careful measurements in order to acquire information for designing a new system or to improve an existing one. For example, if the system has any complexity, the question of absolute and relative stability crops up. It is then necessary to establish possible sources of instability. One method is to measure the phase shifts associated with steady-state frequency inputs introduced by the components and subsystems or to perform some other test-yielding equivalent information. Again, in this situation, the determination of system parameters can prove to be very useful.

A related problem is a search for sources of difficulties that impair the performance of a system. It may be an unexpected instability that appears when the system is turned on, or that occurs intermittently. An undesirable vibration may be set up in the system because of mechanical construction or some other type of oscillation may occur that must be removed. A vibrating source like a gasoline engine supplying power to the system may adversely affect the performance of a system. It may be necessary to find the connecting links in order to effect the requisite isolation. Vibratory problems occur in aerodynamic applications, such as wing flutter or blade flutter in turbines. An investigation generally requires both theoretical and experimental phases in order that its objectives be accomplished. Trouble-shooting and fault-finding are important uses of dynamic measurements.

If we are concerned with an operational system, then implicitly or explicitly it is assumed that there are functional relationships between all of the input quantities (interfering as well as desired) and the output quantities. For a well-behaved system the relationships must be unique. In the simplest situations the relationships may be expressed as differential equations characterized by parameters such as differential-equation coefficients, sensitivities, time constants, undamped natural frequencies, and damping ratios. When the test program is set up to yield the functional relationships and the associated parameters, then the objectives

given by the list in the first paragraph are also obtained. The one exception is fault-finding when the source of the fault is extraneous to the proper performance of the system. From this point of view the various reasons for testing physical systems are specialized aspects of the more general problem of determining the functional relationships and the parameters associated with the system performance. The discussion that follows lays much stress on this general objective.

Preparation Before Testing

Two limiting situations may be recognized when it is necessary to test a system for its performance characteristics. On the one hand, the system may be one whose operation is well understood both theoretically and practically. The test program is designed to supply values for the parameters in functional relationships that are known and expressed in usable form. At the other extreme is the so-called black box problem where little or nothing is known other than the more obvious input and output quantities. The black box problem is probably an academic exercise. The more usual situation lies somewhere between these extremes.

It is to be noted that in the first instance we are dealing with a system about which a considerable amount of information already exists. The information may be derived solely from theory, or it may include results from prior experiments on similar systems. The black box problem is a situation of essentially total ignorance. If the decision is to forego any theoretical preparation, then it is necessary to plan an extensive test program to supply the information initially available in the first situation. For example, it is necessary to determine if the input quantities interact with the system parameters and to ascertain the effect of altering the input-quantity levels. It is necessary to account for all of the significant input quantities and to maintain the modifying and interfering inputs at the proper values. Most often much preparation is required before the dynamic test program can be planned and executed. In the absence of any guidance, the tests may be redundant or produce trivia.

The value of laying down the theoretical base for the performance of the system to be tested is to effect an economy in the over-all program. The very process of developing the theory will generally provide the clues as to the quantities that should be measured. The depth and profundity of the theoretical analysis will depend upon the particular situation. The dynamic measurements may be used to provide a measure of the validity of the theory. When the results prove to be more inconsistent than can be tolerated, the underlying theory should be re-examined. The argument presented in this paragraph is directed toward giving greater recognition to the importance and the value of developing a good theoretical basis before carrying out the tests. In this connection it cannot be overemphasized that the assumptions underlying the theory should be written down clearly and distinctly, since difficulties with the theory often stem from an unwarranted assumption. There is nothing wrong with making all the assumptions necessary to produce the most simple theory to describe a system. It is necessary, however, to be able to ascertain that the assumptions fit the situation. With a well-developed theoretical background, the dynamic measurements of a particular physical system may be obtained with the greatest economy of effort and time and with the maximum efficiency.

Criteria in Choosing the Test Method

Once the theory is well in hand the quantities to be measured may be determined, at least tentatively. The data and results may indicate the existence of deficiencies in the theory. The next task is to choose the test methods. These will depend on the theory, the particular system, the technique for executing the tests, and the method for evaluating and interpreting the data.

Conversely, the techniques for interpreting the data will depend upon the choice of the test procedure. The two go hand in hand and should be planned together. It is worth the effort to cull all the quantities that can be measured quasistatically. These will include the measurements needed to determine sensitivities, usable range of variables, levels of uncertainties, and the system parameters that are available from static tests. It may very well happen that many of these quantities can be determined from dynamic measurements. The static measurements will provide check values on the results of the dynamic measurements.

When the background has been completed, there is generally a choice among several methods of procedures for carrying out the dynamic measurements as well as for evaluating and interpreting the results. The criteria for choosing one method in preference to others include the following: (a) Will the method test the system within the conditions imposed by the available theory? (b) Is the method feasible under the normal operating conditions of the system? (c) Can the method yield reliable results? (d) What is the character of the information made available by the method? (e) What requirements are imposed on the system in order to carry out the tests? (Does it require the system to be taken out of normal operation like taking a process plant off stream?) (f) How much auxiliary equipment is necessary? (g) How much does the method cost in money and time to achieve the necessary accuracy? (h) What is the caliber of the personnel and the level of training in order to perform the tests and interpret the tests effectively?

There is in the interpretation of the results an additional factor of considerable significance that is usually buried in the theoretical basis of dynamic measurements. A survey of the test procedures shows that it is generally assumed that physical systems have linear-system performance in some sense. Relatively few references exist on the theory of making dynamic measurements of nonlinear systems. The history of engineering practice shows it to have been almost exclusively concerned with linear systems because of the relative simplicity of the associated mathematical apparatus. The theory of nonlinear systems is still being developed and has none of the maturity and currency of linear-system theory. Static measurements of the performance of non-linear systems can be and are performed. However, without a well-developed theoretical background, it is hard to see how to develop a program for making dynamic measurements with the generality that is associated with linear systems. It is worth while to examine for the moment the generalized concept of the linear system.

The Concept of the Linear System

Distinctions have been drawn in the literature between linear and nonlinear systems. The emphasis has been so heavily laid as to imply that nonlinear systems represent a revolutionary new technological approach. History shows the contrary, that many physical systems, demonstrably nonlinear, have been studied, designed, built, tested, and successfully operated employing only linear-system theory.

The expression "linear system" has come to mean a system that can be associated with a linear differential equation with constant coefficients. To the mathematician the concept of linearity is applicable to more general types of equations. However, the definition just stated is the one usually implied by engineers. The advantages of linear systems are that the response to an arbitrary function may be readily calculated from well-known, commonly available functions, and that the response to a sequence of arbitrary inputs may be obtained by adding the responses to each of the individual arbitrary inputs. The second property is called the principle of superposition. A system that cannot be associated with a linear equation is a nonlinear system. Linear-system theory is used extensively because general solutions are more easily obtained than with nonlinear systems. In fact, general solutions are not usually available for most nonlinear systems, except in formalistic terms.

For physical systems, the coefficients of the differential equation are usually functional relationships of the properties of the system components. The linear differential equation is often derived by associating physical laws or principles with the operation of the system. It is well known that no physical system is truly linear for all values of its inputs and outputs. From an engineering standpoint it is necessary to establish tolerances and limits of the input and output ranges. The concept of system linearity is valid only within the specified tolerances and limits. In this sense, the use of linear differential equations with constant coefficients is the limiting case for a more general relationship that may not be mathematically expressed. The operation of obtaining the simpler linear equation is defined as linearizing the general relationship.

It is possible to distinguish several levels in the application of the linear-system concept. The most elementary level is a system that is associated with a linear differential equation whose coefficients remain constant and invariant for all values of the system inputs and outputs. A more general system is one that is associated with a linear differential equation with coefficients that may vary with the operating level of the inputs and outputs. A still more difficult situation is encountered by systems whose internal components operate nonlinearly but whose over-all system performance may be represented by a linear equation. These do not exhaust all possibilities.

Simple measuring instruments are often required to have constant static and dynamic characteristics. Galvanometers and galvanometric devices like voltmeters and ammeters, thermometers, pressure pickups, and mechanical position transducers are typical simple measuring instruments with linear characteristics. Great pains are taken to insure linear performance within the operating range and to minimize the effects of unwanted inputs. The usefulness of these instruments depends upon the constancy of the instrument properties. Dynamically, the instrument performance is representable by linear differential equations with constant invariant coefficients under all conditions of operation and throughout the entire usable input range.

Other systems and measuring instruments do not have characteristics that are invariant with the operating level. When the variations are relatively small it may be feasible to make the characteristics constant by restricting the range or by introducing corrective elements like nonlinear springs. Generally this is neither a useful nor a practical procedure. For example, the performance of an airplane flying at one altitude and attitude under one set of load distribution and trim conditions is considerably different from what the performance will be for a different altitude or for different operating conditions. Similarly, the operating characteristics of a heat exchanger that transfers thermal energy from one fluid to another are determined by the temperature differential between the fluids, the rate of flow, and the character of flow, that is, laminar or turbulent. Likewise, electronic-amplifier performance depends on the supply voltages and the level of the input and output signals. There are many other examples, but these illustrate the kind of variability intended.

The performance of systems with variable characteristics is nonlinear. However, the systems just described are generally operated at some fixed level that may be slowly changing to another level, as in the aircraft example. It is desirable to know the response of the system to small deviations from the reference level. In particular, it is necessary to determine the stability of the system, that is, whether it will return to the fixed operating level or seek another. Linear-system theory is applicable when it can be established that the performance for small deviations can

be predicted within acceptable tolerances by the theory. It may be sufficient to carry out the demonstration mathematically and very often this is the situation. In other instances it is necessary to perform experiments in order to gain the required information.

The performance of the linearized system is associated with a linear differential-equation form and with a table relating the equation coefficients to the operating conditions. The coefficient table may be generated from theory when it exists. It is then only necessary to evaluate the system performance for a few reference operating conditions in order to provide quantitative data for the complete coefficient table. In the absence of an acceptable theory it is necessary to carry out an extensive test program for all operating conditions. A general test program is both expensive and time consuming, so that it may not be feasible to investigate the performance outside of the normal operating conditions. This again shows the value of establishing a useful theory for the performance of a system before testing it.

No distinction has been drawn between closed-chain and open-chain systems in the previous discussion. It is to be observed, however, that the essence of closed-chain system performance is the use of deviations to actuate the system. In its crudest aspects a closed-chain system may be represented as in Fig. 1. The input is compared with the feedback to produce a deviation represented as the output-correction quantity. The system is so organized that the output is brought into consonance with the input. It is desirable that the output correction quantity should be small. A
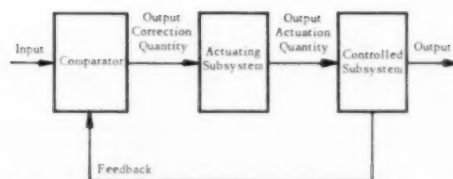
Fig. 1 A simple representation of a closed-chain system

well-behaved system operates to reduce the deviation to zero and, in general, it is responding to small deviations from reference levels. This is a situation wherein it is to be expected that linear-system theory should serve very well to describe the closed-chain system performance, even when some of the components are severely nonlinear. It is necessary, however, to make sure that the deviations between the output and input are indeed small before applying the theory. It is not necessary that the coefficients remain invariant for all ranges of the input. The more general method of using a linear equation form and a function table relating coefficients to operating conditions may be needed. The more general method may be called the controlled point analysis for closed-chain systems. For a well-behaved closed-chain system the output usually takes on the proper values as required by the input. The linearized theory is used to determine how well the system performs.

Summary of Equation Forms Associated With a Multicoupled Operating System
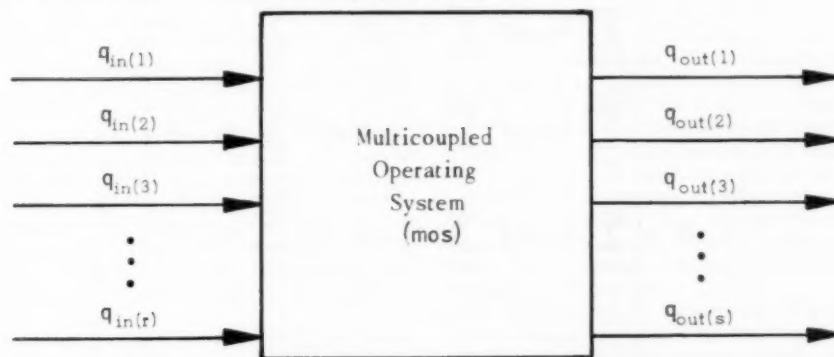
a) Multicoupled operating system with r inputs and s outputs

For the linearized system shown in the figure of section a, the s outputs are related to the r inputs by a set of equations as follows:

$$\sum_{i=0}^{n_{11}} a_{i(11)} \frac{d^i}{dt^i} q_{out(1)} + \sum_{i=0}^{n_{21}} a_{i(21)} \frac{d^i}{dt^i} q_{out(2)} + \cdots \left.\rule{0pt}{5em}\right\}$$

$$= \sum_{k=0}^{m_{11}} b_{k(11)} \frac{d^k}{dt^k} q_{in(1)} + \sum_{k=0}^{m_{21}} b_{k(21)} \frac{d^k}{dt^k} q_{in(2)} + \cdots \qquad (1a)$$

$$\sum_{i=0}^{n_{12}} a_{i(12)} \frac{d^i}{dt^i} q_{out(1)} + \sum_{i=0}^{n_{22}} a_{i(22)} \frac{d^i}{dt^i} q_{out(2)} + \cdots$$

$$\left. = \sum_{k=0}^{m_{12}} b_{k(12)} \frac{d^k}{dt^k} q_{in(1)} + \sum_{k=0}^{m_{22}} b_{k(22)} \frac{d^k}{dt^k} q_{in(2)} + \cdots \right\} \tag{1b}$$

etc.

In general there are as many equations in the set as there are output quantities. When there are s output quantities there are s equations. The set may be written compactly by introducing operational notation. Let

$$p = \frac{d}{dt} \quad ; \quad p^2 = \frac{d^2}{dt^2} \quad ; \quad p^i = \frac{d^i}{dt^i} \tag{2}$$

The equation set may be expressed as

$$\sum_{c=1}^{s} \left\{ \sum_{i=0}^{n_{ce}} a_{i(ce)} p^i \right\} q_{out(c)} = \sum_{d=1}^{r} \left\{ \sum_{k=0}^{m_{de}} b_{k(de)} p^k \right\} q_{in(d)} \tag{3}$$

where $e = 1, 2, \ldots, s$

b) *Equation set associated with multicoupled operating system*

Form the characteristic determinant

$$\Delta = \begin{vmatrix} \sum_{i=0}^{n_{11}} a_{i(11)} p^i & \sum_{i=0}^{n_{21}} a_{i(21)} p^i & \cdots & \sum_{i=0}^{n_{s1}} a_{i(s1)} p^i \\ \sum_{i=0}^{n_{12}} a_{i(12)} p^i & \sum_{i=0}^{n_{22}} a_{i(22)} p^i & \cdots & \sum_{i=0}^{n_{s2}} a_{i(s2)} p^i \\ \sum_{i=0}^{n_{13}} a_{i(13)} p^i & \sum_{i=0}^{n_{23}} a_{i(23)} p^i & \cdots & \sum_{i=0}^{n_{s3}} a_{i(s3)} p^i \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \sum_{i=0}^{n_{1s}} a_{i(1s)} p^i & \sum_{i=0}^{n_{2s}} a_{i(2s)} p^i & \cdots & \sum_{i=0}^{n_{ss}} a_{i(ss)} p^i \end{vmatrix} \tag{4}$$

Form the derived determinants by replacing the cth column of Δ by the operator coefficients of the dth input in the equation set (3)

$$\Delta_{cd} = \begin{vmatrix} \sum_{i=0}^{n_{11}} a_{i(11)} p^i & \cdots & \sum_{k=0}^{m_{d1}} b_{k(d1)} p^k & \cdots & \sum_{i=0}^{n_{s1}} a_{i(s1)} p^i \\ \sum_{i=0}^{n_{12}} a_{i(12)} p^i & \cdots & \sum_{k=0}^{m_{d2}} b_{k(d2)} p^k & \cdot & \sum_{i=0}^{n_{s2}} a_{i(s2)} p^i \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \end{vmatrix} \tag{5}$$

From Cramer's rule, the equation set (3) may be written with only one output quantity in each equation

$$\Delta q_{out(c)} = \sum_{d=1}^{r} \Delta_{cd} q_{in(d)} \qquad (6)$$

where $c = 1, 2, \ldots, s$.

c) Determinantal form of the performance equations

Each equation in (6) represents a unique functional relationship between a particular output and all the inputs Define

$$[FF]_{(mos)[d;c]} = \frac{\Delta_{cd}}{\Delta} = \text{multicoupled operating system performance function for the } d^{th} \text{ input and } c^{th} \text{ output} \qquad (7)$$

The equation set relating the outputs to the inputs of the multicoupled operating system may also be expressed in performance function form as

$$q_{out(c)} = \sum_{d=1}^{r} [PF]_{(mos)[d;c]} q_{in(d)} \qquad (8)$$

where $c = 1, 2, \ldots, s$

d) Performance functions associated with a multicoupled operating system

Multicoupled Operating System Concepts

The most general situation that will be encountered is a multi-coupled operating system with several inputs and several outputs. A symbolic representation of the system is shown in the figure of the accompanying summary of equation forms. The inputs include the desired input quantities, the modifying inputs, and the interfering inputs as defined in vol. 1 of "Instrument Engineering" (1).[2] The outputs are usually taken as the desired outputs, although no restriction is implied. The linearized equations associated with the performance of a generalized multicoupled operating system are given in the summary in several useful forms. Inasmuch as there are several inputs and outputs, it is necessary to provide a set of equations. In general there should be as many equations as there are output quantities. If there are fewer equations, the set is indeterminate; if there are more, it is redundant. It may not always be possible to satisfy the condition on the number of equations, in which case the situation requires special treatment, particularly when redundancy exists. For the general treatment shown in the summary, it is assumed that the number of equations in the set is equal to the number o output quantities.

The set of relationships given by Equation [1] is intended to be the general form for a set of integro-differential equations between the outputs and the inputs. Each term shown in each equation is a summation of the derivatives and integrals of one quantity, either an input or an output. The output quantity coefficients are indicated by the letter *a* with a subscript associating it with the derivative, the output quantity, and the equation. The input quantity coefficients are indicated by the letter *b* with a comparable subscript. The coefficients may be positive, negative, or zero. The technique employed in Equation [1] shows all the terms in all the equations but is very cumbersome for the generalized situation. It is more convenient to employ an operational notation as indicated by Equation [2], where the derivative is replaced by the letter *p* to a power. Positive powers indicate derivatives and negative powers mean integrals. The equation set may be expressed compactly by Equation [3] using the operational notation and double summations. Equation [3] may be considered as a typical form of the entire set of equations associated with the performance of the multicoupled operating system.

There are well-known mathematical procedures for manipulat-

[2] Numbers in parentheses refer to Bibliography at end of paper.

ing the equation set into a form where each equation has only one output. From Cramer's rule, for example, the new equation set may be expressed as Equation [6] and there are just as many equations in the new set as in the old. Two special symbols have been employed in Equation [6]. The delta symbol without subscript is defined by Equation [4] as the determinant formed from the left-hand-side summation coefficients in Equation [3]. The notation is a little more general than usually encountered, but the technique is the same. In Equation [4], each element of the determinant is a summation. When Equation [4] is expanded and set equal to zero, it forms the characteristic equation associated with the equation set. The roots of the characteristic equation determine the modes of the multicoupled operating system (1, 2).

The delta symbol with subscripts is defined by Equation [5]. It is the characteristic equation determinant in which one column is replaced by the summation coefficients of one of the input quantities on the right-hand side of the equation set. The first subscript indicates the output quantity whose coefficients are replaced. The second subscript indicates the input quantity whose coefficients are substituted.

The Equation Set [6] is interpreted as a new set of integro-differential equations written in operational form. Each output quantity is acted upon by the same operator, namely, the characteristic equation polynomial. The right-hand side is a summation of input quantities, each modified by a unique operator as defined by Equation [5]. The Equation Set [6] shows that every output has the same set of transient modes but is influenced differently from the other outputs by the input quantities. These are well-known results from the theory of vibrations.

It is also useful to express the equation set in another way, as given by Equation [8]. A new operator is defined by Equation [7] as the performance function relating a single input to a single output. The concept of the performance function is developed in vol. 1 of "Instrument Engineering" (1). The Laplace transform representation of the performance function may be identified as the transfer function, but the expression given by Equation [7] is not a Laplace transform. In Equation [8], an output quantity is related to the inputs by a summation of performance functions. The subscript notation is employed to relate a particular performance function to the associated input and output quantities.

The development in the summary leading from the differential equation set to the performance function equation set is possible because of the linear character of the equations. In Equation

[8] it may be seen that each output is given by the sum of the effects of the inputs. The equations suggest that the influence of each input may be determined by setting all but one equal to zero.

It is sometimes useful to factor the characteristic equation polynomial and all of the derived determinants. The roots of the determinants may be associated with each performance function as a method of characterizing the operator. In the simple situation where there is only one input and one output quantity, the equation set reduces to one expression

$$\sum_{i=0}^{n} a_i \frac{d^i q_{out}}{dt^i} = \sum_{k=0}^{m} b_k \frac{d^k q_{in}}{dt^k} \dots \dots [9]$$

or

$$\left\{ \sum_{i=0}^{n} a_i p^i \right\} q_{out} = \left\{ \sum_{k=0}^{m} b_k p^k \right\} q_{in} \dots \dots [10]$$

The characteristic determinant has only one element

$$\Delta = \left| \sum_{i=0}^{n} a_i p^i \right| = \sum_{i=0}^{n} a_i p^i \dots \dots [11]$$

There is only one derived determinant and it has a single element

$$\Delta_{11} = \left| \sum_{k=0}^{m} b_k p^k \right| = \sum_{k=0}^{m} b_k p^k \dots \dots [12]$$

The performance function is

$$[PF] = \frac{\Sigma b_k p^k}{\Sigma a_i p^i} \dots \dots [13]$$

The roots of Equation [11] are sometimes called the poles of the performance function and the roots of Equation [12] are its zeros. The pole-zero nomenclature can be extended to the general situation with the addition of appropriate designators if it is desired.

The task of interpreting the dynamic measurements of a multicoupled operating system may be expressed in any of several methods, as indicated by the several ways for setting up the performance equations. One part of the task may be regarded as the problem of determining the form of the performance functions; the rest of the task is to ascertain the parameters. They may be taken as the coefficients on both the left and right-hand sides of the differential equation set as given by Equation [1] or in the altered form as Equation [6]. It is equally useful to express the parameters as the coefficients of the terms in the performance functions, or again, as the roots of the numerator and denominator polynomials of the performance functions. The various methods for describing the system parameters may be employed in dimensional or nondimensional form as suits the occasion.

Test Procedures

When the objective of the test program is to determine the system parameters, it is intended to find numerical values for these quantities in one or more of the forms indicated in the summary. The dynamic measurements show only how the input and output quantities vary as a function of time. The measurements are interpreted by evaluating the parameters from the time histories. The techniques and procedures that are necessary to get the time histories are significant in evaluating the limitations of the interpretation procedures. It is not intended to discuss the techniques of dynamic measurements in this series of papers, but the importance of this phase cannot be overstressed. It will be necessary to consider the influence of the processes of making measurements on their interpretation. Fig. 2 is an idealized representa-

tion of a functional diagram for the dynamic measurement of the performance of a multicoupled operating system with three inputs and three outputs. Generally the input and output quantities cannot be recorded as such. It is usually necessary to provide instruments, shown as input and output quantity receivers. Each receiver converts a quantity to be measured into a signal suitable for actuating one element of a multichannel recorder. The recorded time history for a given quantity is generally several stages removed from the quantity itself. This is a condition that has considerable significance in establishing the limits of the interpretation technique. Part of the problem of making and recording dynamic measurements is the selection of receivers and recorders that do not introduce dynamic characteristics of their own. This aspect will not be taken up in the interpretation of dynamic measurements.

The various procedures for testing the dynamic properties of physical systems may be classified by the way the time histories are generated. These are (a) transient studies, (b) steady-state frequency studies, (c) arbitrary and random-input studies. When the data may be used essentially as measured to obtain the system parameters, the process may be described as direct interpretation. There are other methods whereby the data must be processed into new forms before the system parameters are evaluated. These are the methods of indirect interpretation. The random input studies usually require processing before the information can be used for system parameter evaluation. The information from the other two dynamic measurement procedures may be employed either directly or indirectly.

Transients

The solution of the multicoupled operating system to a variation of one of the inputs may be regarded as consisting of two parts—the transient solution and the forced solution. The time variation of each of the outputs may be written as

$$q_{out(c)} = q_{out(c)(ts)} + q_{out(c)(fs)} \dots \dots [14]$$

where

$q_{out(c)(ts)}$ = transient-solution part of the cth output
$q_{out(c)(fs)}$ = forced-solution part of the cth output
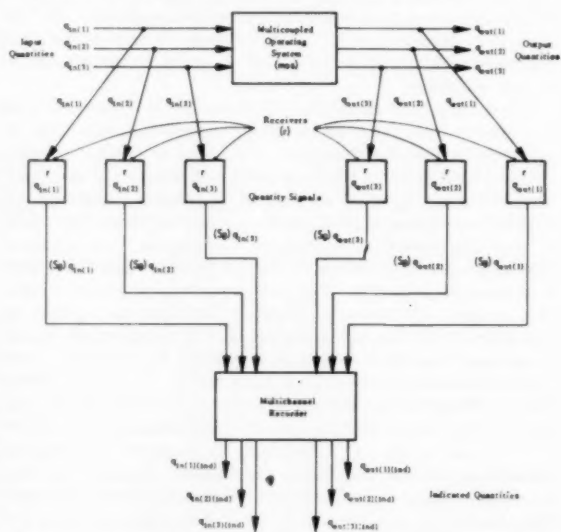
Fig. 2 Functional diagram for the dynamic measurements of a multicoupled operating system with three input and three output quantities

The transient-solution part can have the form

$$q_{out(c)(ta)} = \sum_i A_{i(c)} e^{z_i t} \dots \dots \dots [15]$$

where

z_i is a root of the characteristic equation

$$\Delta = 0 \dots \dots \dots \dots \dots [16]$$

It is assumed in Equation [15] that all the characteristic equation roots z_i are distinct. When there are multiple roots, the expression for the transient solution becomes somewhat more complicated. The procedure to be used for multiple roots is explained in many places, see, for example (1, 2). The coefficients $A_{i(c)}$ are complex numbers that depend upon the initial conditions of the inputs and outputs and upon the functional form of the input variation.

In general, the characteristic equation roots are complex quantities

$$z_i = x_i + j y_i \dots \dots \dots \dots [17]$$

From solutions of first-order equations, when the root is real

$$z_i = x_i = -\frac{1}{\tau_i} \dots \dots \dots \dots [18]$$

where τ_i is the characteristic time. From solutions of second-order equations when the root is complex

$$z_i = -\zeta_i \omega_{n(i)} + j\omega_{n(i)}(1 - \zeta_i^2)^{1/2} = \omega_{n(i)} e^{j(\pi - A_{\zeta i})} \dots [19]$$

$$x_i = -\zeta_i \omega_{n(i)} = \omega_{n(i)} \cos(\pi - A_{\zeta i}) \dots \dots [20]$$

$$y_i = \omega_{n(i)}(1 - \zeta_i^2)^{1/2} = \omega_{n(i)} \sin(\pi - A_{\zeta i}) \dots [21]$$

$$\omega_{n(i)} = \frac{2\pi}{T_{n(i)}} = \text{angular undamped natural frequency} \ . [22]$$

ζ_i = damping ratio, $A_{\zeta i} = \cos^{-1} \zeta_i$ = angle of damping ratio
$$\dots \dots [23]$$

The roots may be plotted on the root-complex plane as shown in Figs. 3 and 4. Direct test procedures using transient studies attempt to determine the characteristic equation roots and the coefficients of Equation Set [15] from the transient responses to known inputs.
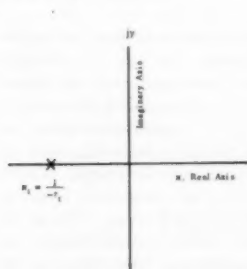
Fig. 3 Plot of a real characteristic equation root
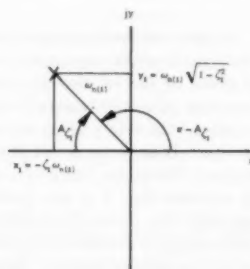
Fig. 4 Plot of a complex characteristic equation root

Steady-State Frequency Response

The steady-state solution of the multicoupled operating system is easily calculated when the input is a constant-amplitude, constant-frequency sinusoidal variation. To simplify the situation somewhat, let all but one input be constant. Let the dth input quantity be

$$q_{in(d)} = q_{in(d)(a)} e^{j\omega t} \dots \dots \dots \dots [24]$$

where

$q_{in(d)(a)}$ = amplitude of the dth input

ω = angular forcing frequency

$j = \sqrt{-1}$

The steady-state solutions are calculated from Equation Set [3] when the operator p is replaced by $j\omega$

$$\sum_{c=1}^{s} \left\{ \sum_{i=0}^{n_{ce}} a_{i(ce)}(j\omega)^i \right\} q_{out(c)} = \left\{ \sum_{k=0}^{m_{de}} b_{k(de)}(j\omega)^k \right\} q_{in(d)} \dots [25]$$

where $e = 1, 2, \dots, s$. Solving by Cramer's rule, the equation set becomes

$$q_{out(c)} = \frac{\Delta_{cd}}{\Delta} q_{in(d)} = [PF]_{(mos)[d; c]}(j\omega) q_{in(d)} \dots \dots [26]$$

where

Δ = characteristic equation determinant defined by Equation [4] with p replaced by $j\omega$ $\dots \dots \dots \dots [27]$

Δ_{cd} = determinant defined in Equation [27] with the cth column replaced by the bracketed coefficient on the right-hand side of Equation [25] $\dots \dots \dots \dots [28]$

$[PF]_{(mos)[d; c]}(j\omega)$ = frequency form of the performance function defined by Equation [7] when p is replaced by $j\omega$ $\dots \dots [29]$

The frequency function defined by Equation [29] has two parts, an amplitude ratio function and a dynamic response angle (phase angle) function.

The frequency responses may be calculated by the breakpoint method. As explained in several references, the break points are functionally related to the roots of Equations [27] and [28]. First-order factor break points are defined as

$$(bp)_{(1st\,ord)} = \frac{1}{\tau} \dots \dots \dots \dots [30]$$

when the forcing frequency is in radians per second.

Second-order factor break points are equal to ω_n. The appropriate damping ratio determines the shape of the second-order frequency curve

$$(bp)_{(2nd\,ord)} = \omega_n \dots \dots \dots \dots [31]$$

Direct test procedures using steady-state frequency responses attempt to determine the break points from experimentally derived amplitude ratio and dynamic response angle curves. The differential equation coefficients are computed by inverting the process for finding the roots.

Arbitrary and Random Input Studies

The response of the cth output quantity to a unit impulse for the dth input is defined as the weighting function

$$[WF]_{(mos)[d; c]}(t) = c\text{th output for unit impulse at } d\text{th input} \ . \ . [32]$$

= multicoupled operating system weighting function for cth output and dth input.

From the principle of superposition the cth output due to variations of the dth input is calculated from the convolution integral (1)

$$q_{out(c)} = \int_{-\infty}^{t} [WF]_{(mos)[d; c]}(t - t_i) q_{in(d)}(t_i) dt_i \dots [33]$$

By further application of the principle of superposition, the cth output may be related to all the inputs

$$q_{\text{out}(c)} = \sum_{d=1}^{r} \int_{-\infty}^{t} [WF]_{(\text{mos})[d;\ c]}(t - t_i)q_{\text{in}(d)}t_i dt_i \dots [34]$$

In passing it is to be noted that the Fourier transform of the weighting function is the frequency form of the performance function defined in Equation [26]

$$[FT][WF]_{(\text{mos})[d;\ c]}(t) = [PF]_{(\text{mos})[d;\ c]}(j\omega) \dots [35]$$

Techniques exist for deducing the weighting function when an arbitrary input is applied. Other techniques have been developed for determining the function when the input is a randomly fluctuating quantity.

Procedures based on random inputs require the time histories be recorded of all the inputs and outputs for sufficiently long periods of time. The weighting functions are deduced from the records with the aid of statistical procedures to facilitate the process.

Experimental Limitations

If it is possible to establish an objective basis for choosing the test procedure and the method for interpreting the data, it will probably be developed from a consideration of the experimental limitations inherent in dynamic measurements. There are at least five sources of experimental limitations:

1 Limitations as to the quantities that are capable of dynamic measurement.

2 Limitations imposed by the theory.

3 Limitations in the capabilities of the system, such as saturation and uncertainty.

4 Limitations in the static and dynamic characteristics of the measuring instruments.

5 Limitations on the signal originating in the corruptions introduced by the measuring process (sometimes called noise).

Referring once again to Fig. 2, it is to be noted that the recorded data are generally several steps removed from the quantity being measured. The details in the process of getting a record of a quantity are the proper subject for instrument engineering. Fortunately, it is possible to characterize the measuring techniques in a quantitative manner without delving extensively into the instrument engineering field.

The first of the limitations mentioned in the list is a recognition that some quantities cannot be measured for physical or technological reasons. This is the condition, for example, in many chemical plants where the composition of the raw materials or the end products must be analyzed from a sample in the laboratory. In most physical systems the direct measurement of derivatives higher than the second order is not possible with available instruments.

Where linearized theory is employed, the input and output quantity variations must be restricted to that range for which the theory is applicable. This establishes an upper limit on the magnitudes of the signals representing the variations. Similar limits exist when system components are capable of saturation. The possibility of saturation requires care in the selection of the test function applied to the inputs. If a system is subjected to a step function input, for example, the input quantity must suffer a discontinuity as it goes from one level to another. In many instances it is impossible to apply a true step function; a very rapid change is substituted. The change in level must occur so quickly that the system cannot effectively respond before the change is completed. In these circumstances the system outputs are required to attain their new levels as quickly as possible. It

often happens that some components are required to operate at levels exceeding their capabilities and suffer saturation. These conditions must be avoided to obtain usable dynamic measurements. Saturation may also occur when large pulses of very short duration are applied.

At the other extremity, there are limits to the minimum values for the measured quantities. In a sense, this is a limitation of the measuring elements, but it may be considered a property of the system itself to distinguish it from the limitations of the measuring instruments. Experience shows that there always seems to be a level of uncertainty[3] in the value of the input and output quantities of physical systems (3).

In addition to the inherent uncertainty associated with the measured quantities, there are uncertainties introduced by the measurement technique. Some of these are due to the measuring instruments and some may be attributed to the environment in which the measurements take place. The additional uncertainties arising in the measuring process may be called the corruption of the signal; sometimes, it is also referred to as noise. The sources of the signal corruption may be in the coupling between the measured quantity and the measuring instrument, in the instrument itself, or from extraneous sources that cannot be suppressed.
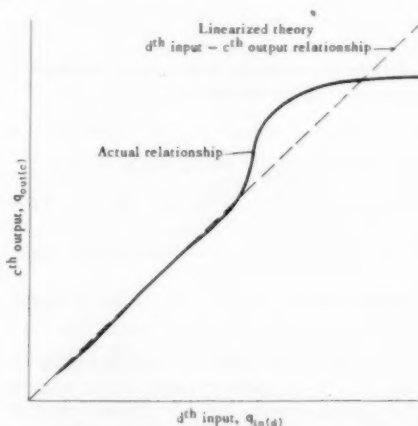
Fig. 5 Quasistatic relationship between cth output and dth input of a multicoupled operating system (example 1)

A notion of the meaning of the experimental limitations may be obtained from the accompanying figures. In Fig. 5 an example is given of the relationship between one input and one output quantity of a multicoupled system. A dashed line is shown to represent the relationship required by the hypothetical linearized theory. It is to be noted that the actual relationship does not begin at the origin, but at some distance from the origin. This is to indicate that it is not possible to obtain consistent measurements in the region of small-input quantity values. The actual relationship is drawn reasonably close to the linearized theory line until it suddenly departs, first by increasing faster than for the linear relationship and then by leveling off. The variation from the linearized theory may be seen more clearly in Fig. 6 where the sensitivity function is plotted. The sensitivity for the given input and output is defined in the figure as the partial derivative of the output with respect to the input. For small values of the input, there is no function because this is the uncertainty range. The limit to the linearized theory range is located only for illustra-

---

[3] The developments of information theory also lead to the opinion that there are inherent uncertainties in the input and output quantities.
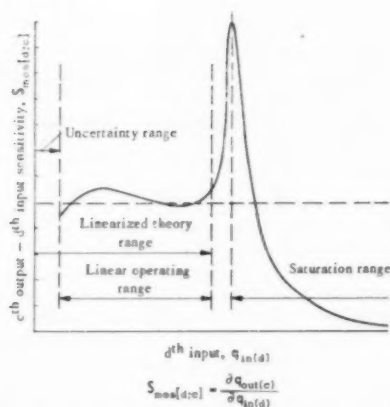
Fig. 6   Variation of *c*th output—*d*th input multicoupled operating system sensitivity for example 1 (see Fig. 5)
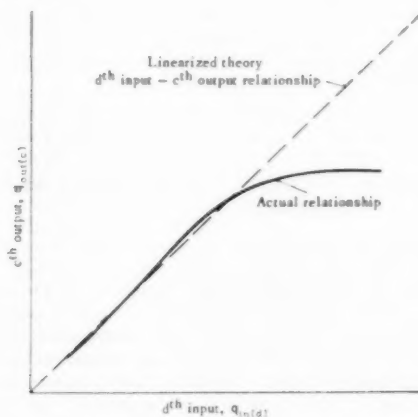
Fig. 8   Variation of the *c*th output—*d*th input multicoupled operating system sensitivity for example 2 (see Fig. 7)

Fig. 7   Quasistatic relationship between *c*th output and *d*th input of a multicoupled operating system (example 2)
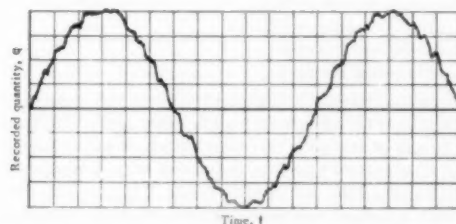
Fig. 9   Example of a recorded quantity showing the presence of uncertainties and signal corruptions

tive reasons. It was set by establishing a 10 per cent tolerance limit in the variation of the sensitivity. The saturation range is assumed to exist after the peak value in the sensitivity. In this example the linear operating range is located between the uncertainty range as a lower limit and the linearized theory range as an upper limit. Another example is shown in Figs. 7 and 8 where the linearized operating range lies between the uncertainty range and the saturation range.

The examples illustrated by the figures are not meant to be ex-

haustive. They show how it is possible to define quantitative bounds on the range of the data representing the measured quantities. The range is limited between the upper limits imposed by theory, by system capabilities, and by the measuring system capabilities, and the lower limits established by the uncertainties in the measured quantities and the corruptions of the signal. Inasmuch as these concepts are amenable to quantitative description, they may be used to provide measures that can be employed in planning and evaluating a test program. It should be possible to employ the measures to develop objective and quantitative statements for which the criteria are given in the foregoing.

An illustration of a recorded quantity showing the presence of uncertainties and signal corruptions is given by Fig. 9. The original quantity was a pure sinusoidal variation. It shows the difficulty of making precise determinations of amplitude and phase. The example represents a good record as compared with many that have been used.

Part II—Methods for Testing Physical Systems and for Interpreting Dynamic Measurements

Part II is the second of a series of papers on interpreting dynamic measurements. A scheme is proposed to classify the testing and interpretation of physical systems into six categories:

1   Derivatives method.
2   Steady-state frequency method.
3   Standard curves method.
4   Indirect frequency method.

5   Random input method.
6   Combination method.

The basis for each method is developed together with a discussion of techniques and limitations.

Introduction

THIS is the second part in a series of papers on the theme of in-

terpreting dynamic measurements. The background, aims, and objectives have been presented in the first part. It has been asserted that it is usually necessary to attribute a linearized character to physical systems subjected to dynamic measurements when the results are to have general applicability. The linearization condition is not as restrictive as might be suggested by superficial examination. It is shown in the previous part how linearization procedures may be applied to a large class of systems including many that are really nonlinear. The most general situation is the multicoupled system with many inputs and outputs. The mathematical background associated with the multicoupled system is set down in the previous part, and should be used for reference in connection with the material presented below.

The existing methods for testing and interpreting the performance of physical systems may be classified according to the following scheme:

1 Derivatives method
   (a) Measurement of derivatives method
   (b) Calculation of derivatives method
2 Steady-state frequency measurements method
3 Standard curves method
4 Indirect frequency method
5 Random input method
6 Combination method

Dynamic measurement is generally useful only in so far as the results obtained by one method can be used freely with those obtained by any of the other methods. When the dynamic characteristics of a system must be associated with a particular test procedure and a different set of dynamic characteristics are associated with a second test procedure for identical operating conditions, the results are of very limited utility. The assumption that the system is linear within the operating range specified by the test procedure gives the desired universal characteristics to the results. The techniques and procedures discussed under the six methods cited in the outline are developed only for linear systems. It is on this ground that it becomes possible to make a comparison among the methods.

Measurement of Derivatives Method

In the measurements of derivatives method, it is necessary to produce a set of linear differential equations from theoretical considerations relating the input and output quantities. As explained in the first part of this paper, the general mathematical expressions associated with a multicoupled operating system are

$$\sum_{e=1}^{s} \left\{ \sum_{i=0}^{n_{ce}} a_{i(ce)} p^i \right\} q_{\text{out}(c)} = \left\{ \sum_{d=1}^{r} \sum_{k=0}^{m_{de}} b_{k(de)} p^k \right\} q_{\text{in}(d)} \quad [36]$$

where $e = 1, 2, \ldots, s$.

In the set of equations the order of the equations is given; that is, the number of input and output quantities are known, together with the number of derivative terms. While the equation set is written in general terms, in any particular situation it represents a well-understood set of relationships. For example, the linearized equations for the longitudinal motion of an aircraft may be expressed as follows[4] when the long-period motions are ignored

$$\left[\frac{mU}{qS} p - C_{z_\alpha} \right] \frac{w}{U} - \frac{mU}{qS} p\theta = [C_{z_{\delta_e}}]\delta_e \quad [37]$$

$$\left[-\frac{c}{2U} C_{m_\alpha} p - C_{m_\alpha} \right] \frac{w}{U}$$

$$+ \left[\frac{I_y}{qSc} p^2 - \frac{c}{2U} C_{m_q} p \right] \theta = [C_{m_{\delta_e}}]\delta_e \quad [38]$$

[4] From M.I.T. Course 16.15 class notes by E. E. Larrabee.

where

w/U = ratio of vertical velocity to horizontal velocity } output quantities
θ = angle of pitch
δ_e = change in elevator deflection—input quantity
m = mass of aircraft
S = wing area
$q = \frac{1}{2} \rho U^2$
ρ = air density
c = wing chord
U = mean forward velocity
I_y = aircraft moment of inertia about its y-axis
$C_{z_\alpha}, C_{z_{\delta_e}}, C_{m_\alpha}, C_{m_q}, C_{m_{\dot\alpha}}$ = stability derivatives (quantities to be determined)

The purpose of the test program is to determine the coefficients for a given set of operating conditions. The system is instrumented to provide a set of simultaneous records for every input and output quantity as well as for every one of the derivatives that appear in the set of equations. Since at every instant the entire set of equations must be satisfied, the values of the measured quantities may be introduced into the equation set. In this way, for example, at time t_1 the equations are

$$\sum \left\{ \sum a_{i(ce)} p^i q_{\text{out}(c)}(t_1) \right\} = \sum \left\{ \sum b_{k(de)} p^k q_{\text{in}(d)}(t_1) \right\} \quad [39]$$

where

$p^i q_{\text{out}(c)}(t_1)$ = value of the ith derivative of the cth output quantity at time t_1

The expressions represent a set of simultaneous linear algebraic equations with the unknown coefficients $a_{i(ce)}$ and $b_{k(de)}$. However, there are fewer equations than coefficients. The equation set is augmented by using other values of time, t_2, t_3, and so forth, until there are as many linear algebraic equations as coefficients. The relative values of the coefficients are determined by conventional algebraic processes. Absolute values require additional tests to establish the absolute value for at least one of the coefficients.

There is no theoretical reason for choosing one class of inputs rather than another. However, from a practical point of view, the technique for applying this method requires the inputs to provide observable values for the measured quantities. For this reason, inputs are selected that induce significant transients in the response. Furthermore, in the interest of simplicity, all the inputs except one may be held constant (zero is included as a constant). A convenient test input is to apply a step change in this quantity. In other instances, pulses may be more suitable. A pulse is a variation of the input of finite duration that leaves the input unchanged after the pulse is completed.

The necessity of measuring all the derivatives of the variables may require extensive instrumentation. It is feasible to employ the measurements of derivatives method for relatively simple systems with low-order derivatives. When the system requires records of high-order derivatives, it may not be possible to provide the measuring instruments except by installing differentiators. All the sources of experimental limitation mentioned in the first paper influence the calculations. The uncertainties and signal corruptions limit the precision of the values of the derivatives. As a rule, the higher-order derivatives are noisier than the lower-order ones. The linearization approximations also introduce some variability in the results. When the records are extensive enough to permit several evaluations of the coefficients, it will be a rare situation indeed that gives precisely the same values for all the evaluations. The problem of the variation in the coefficient

values may be resolved by statistical theory after the influence of the linearization theory has been removed. The older method from the theory of measurements based on the assumption of a Gaussian distribution of the error may be applied to each co-efficient (1). Newer methods are based on minimization of the mean square error (4, 5) that may be applied either to the in-dividual coefficients or to the equations of the set.

Calculation of Derivatives Method

A more feasible approach than measuring all the inputs and outputs and their derivatives is to measure only the quantities that may be obtained with adequate precision. The remaining quantities are calculated from the measured ones. The equation coefficients may be computed by the procedure outlined for the measurement of derivatives method. If the raw data are differen-tiated numerically or by a differentiating mechanism, there will be the same problems stemming from the uncertainty and noise as for direct measurement of the derivatives. However, it may be possible to smooth the data before further processing. The smoothing often requires considerable skill and judgment to avoid losing information. An obvious method is to draw the best smooth curve by hand.

There are other smoothing methods that employ different tech-niques to evaluate the coefficients. A least-squares method is described by Shinbrot (5) and another by Phillips (6). Shin-brot's presentation is given here. The demonstration is for a system with a single input pulse. The input quantity returns to its reference value. The performance equation has the form of Equation [36]. It is assumed for simplicity only that all the roots of the characteristic equation are distinct. The solution after the pulse is completed is

$$q_{out(e)} = \sum_{i=1}^{n} A_{i(e)} e^{z_i t} \quad \text{for} \quad t > T_1 \ldots\ldots\ldots [40]$$

where T_1 is the duration of the pulse.

The first step is to evaluate the coefficients $A_{i(e)}$ and the ex-ponent factors z_i. The first step is carried out by minimizing the mean square of the error

$$[(E)q_{out(e)}]^2 = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} [q_{out(e)} - q_{out(e)(meas)}]^2 dt \ldots [41]$$

where $q_{out(e)(meas)}$ is the measured output quantity.

When the data are available as a set of measured values, at times $t_1, t_2, \ldots, t_r$, the error square condition may be used

$$(ESC) = [(E)q_{out(o)}]^2 = \sum_{r=1}^{\nu} [q_{out(e)}(t_r) - q_{out(e)(meas)}(t_r)]^2 \ldots [42]$$

The steps in Shinbrot's least-squares method follow:

1   A first approximation to the solution form Equation [40] is found by an independent means. Shinbrot suggests the tech-nique described in the measurement of derivatives method. Let the first approximation be written as

$$q_{out(e)(1)} = \sum_{i=1}^{n} A_{i(e)(1)} e^{z_{i(1)} t} \ldots\ldots\ldots\ldots [43]$$

2   Neglecting higher-order terms, compute the variation

$$\Delta q_{out(e)} = \sum_{i=1}^{n} \left(\frac{\partial q_{out(e)(1)}}{\partial A_{i(e)(1)}} \right) \Delta A_{i(e)} + \sum_{i=1}^{n} \left(\frac{\partial q_{out(e)(1)}}{\partial z_{i(1)}} \right) \Delta z_i \ldots [44]$$

where

$$\Delta q_{out(e)} = q_{out(e)} - q_{out(e)(1)}$$

$$\Delta A_{i(e)} = A_{i(e)} - A_{i(e)(1)}$$

$$\Delta z_i = z_i - z_{i(1)}$$

From Equation [40]

$$\Delta q_{out(e)} = \sum_{i=1}^{n} e^{z_{i(1)} t} [\Delta A_{i(e)} + A_{i(e)(1)} t \Delta z_i] \ldots\ldots [45]$$

3   Define

$$\Delta q_{out(e)(meas)} = q_{out(e)(meas)} - q_{out(e)(1)} \ldots\ldots\ldots [46]$$

$$(ESC) = \sum_{r=1}^{\nu} [\Delta q_{out(e)}(t_r) - \Delta q_{out(e)(meas)}(t_r)]^2 \ldots\ldots [47]$$

4   Compute

$$\frac{\partial (ESC)}{\partial (\Delta A_{i(e)})} = 0; \quad \frac{\partial (ESC)}{\partial (\Delta z_i)} = 0, \quad i = 1, 2, \ldots, n \ldots [48]$$

Solve Equations [48] for the increments $\Delta A_{i(e)}$ and Δz_i. The increments are added to the values of $A_{i(e)(1)}$ and $z_{i(1)}$ to give the second approximation values

$$A_{i(e)(2)} = A_{i(e)(1)} + \Delta A_{i(e)}, \quad z_{i(2)} = z_{i(1)} + \Delta z_i \ldots [49]$$

5   Repeat the preceding steps until two successive iterations give the same values to the desired number of significant figures.

6   Calculate the coefficients $a_{i(e)}$ from the roots z_i by using a method from the theory of equations

$$\left. \begin{aligned} a_{n-1} &= -\sum_{i=1}^{n} z_i \\ a_{n-2} &= \sum_{i,j=1}^{n} z_i z_j, \ldots, \\ a_o &= (-1)^n \prod_{i=1}^{n} z_i \end{aligned} \right\} \ldots\ldots\ldots\ldots [50]$$

7   Calculate the coefficients b_k by deriving the solution for the actual input, taking into account the initial conditions of both the input and output quantities and using the values of the constants $A_{i(e)}$ in Equation [40].

Another approach to the least-squares method is discussed in connection with the random input method.

A very elegant technique to obtain smoothed data is applicable for special situations. The time vector method was developed by R. K. Mueller (7) for determining the stability of aircraft. Further applications are given in References (8, 9, and 10). E. E. Larrabee has made extensive and skillful use of the method to reduce aircraft test records to evaluate stability derivatives. Larrabee's methods are specialized to aircraft applications, where the equations of motion are peculiarly well adapted for the tech-nique. Details may be obtained from the foregoing references.

Steady-State Frequency Measurements Method

The technique of applying a steady-state sinusoidal variation to one input of a system and measuring the steady-state sinusoidal variation in the response of a single output has a long history Under the generic title of frequency response, it has been finding ever more applications (11).

In the extension of the frequency-response procedures to a multicoupled system it is necessary to hold all inputs constant except one. The measurements include the forcing frequency, the amplitude of the input quantity variation, the amplitude of the output quantity variations, and the relative phase of each output with respect to the forcing input. The procedure is repeated for as many frequencies as required to determine the properties of

the system, and for as many inputs as are pertinent to the investigation.

For a given forcing frequency applied to the selected input, expressions for the relationships between the measured quantities are obtained from Equation [25].

$$\sum_{c=1}^{s} \sum_{i=0}^{n_{ce}} a_{i(ce)}(jw_{\nu})^i q_{\text{out}(c)(a)} e^{j\phi}(d-c)$$
$$= \sum_{k=0}^{m_{dc}} b_{k(de)}(jw_{\nu})^k q_{\text{in}(d)(a)} \ldots \ldots [51]$$

$$c = 1, 2, \ldots, s$$

where

w_{ν} = νth value of the forcing frequency in angular measure
$q_{\text{in}(d)(a)}$ = amplitude of dth input quantity
$q_{\text{out}(c)(a)}$ = amplitude of cth output quantity
$\phi_{(d-c)}$ = relative phase of $q_{\text{out}(c)}$ with respect to $q_{\text{in}(d)}$ for the frequency w_{ν}
  = $q_{\text{out}(c)}$ to $q_{\text{in}(d)}$ dynamic-response angle.

Another way to express the measurements is taken from Equation [29]

$$q_{\text{out}(c)(a)} e^{j\phi(d-c)} = [PF]_{(\text{mos})[d; c]}(jw) q_{\text{in}(d)(a)} \ldots \ldots [52]$$

From Equation [52], the performance function relating a particular input to a particular output as a function of frequency is

$$[PF]_{(\text{mos})[d; c]} = \frac{q_{\text{out}(c)(a)}}{q_{\text{in}(d)(a)}} e^{j\phi(d-c)} \ldots \ldots \ldots [53]$$

The magnitude of the frequency performance function is the amplitude ratio

$$|[PF]_{(\text{mos})[d; c]}| = \frac{q_{\text{out}(c)(a)}}{q_{\text{in}(d)(a)}} \ldots \ldots \ldots [54]$$

The argument of the frequency performance function is the dynamic response angle

$$\arg [PF]_{(\text{mos})[d; c]} = \phi_{(d-c)} \ldots \ldots \ldots [55]$$

One method of interpreting the dynamic measurements is to find the coefficients $a_{i(ce)}$ and $b_{k(de)}$ in Equation [51] when the amplitudes, dynamic response angles, and forcing frequencies are known. Trimmer (12) has suggested a method that is an extension of the electrical procedure for determining the resistance, inductance, and capacitance of a simple circuit. In this example the performance equation takes the form

$$\left[L \frac{di}{dt} + \frac{1}{C} \int_0^t i\,dt + Ri \right] = e \ldots \ldots \ldots [56]$$

where e is the applied voltage and i is the measured current. The resistance is measured at zero frequency

$$Ri_0 = e_0 \ldots \ldots \ldots \ldots \ldots \ldots [57]$$

At another frequency Equation [56] becomes

$$\left[Ljw + \frac{1}{Cjw} + R \right] i_a e^{j\phi} = e_a \ldots \ldots \ldots [58]$$

The real part of the equation is

$$\left[-Lw \sin \phi + \frac{\sin \phi}{Cw} + R \right] i_a = e_a \ldots \ldots \ldots [59]$$

The imaginary part is

$$Lw \cos \phi - \frac{\cos \phi}{Cw} = 0 \ldots \ldots \ldots \ldots [60]$$

The values of R, L, and C are obtained by solving the three simultaneous Equations [58], [59], and [60].

Generalizing the procedure to a multicoupled system where the orders of the performance equations are known, the frequency response is measured for half as many frequencies as there are coefficients to be determined. The measurements provide values to be introduced into sets of equations having the form of Equation [51]. The real and imaginary parts of the left-hand and right-hand sides of each equation are equated to each other, respectively. This provides a set of linear simultaneous algebraic equations with the same number of equations as unknowns. Any appropriate mathematical procedure may be used to determine the desired coefficients. When the number of unknowns is greater than four, it may be advisable to use automatic computing machinery.

The more customary practice is to relate the frequency response measurements by the formulations of Equations [53], [54], and [55]. For each frequency the amplitude of the input is divided into that of the output and the relative phase of the output obtained by comparison with the input. The computed quantities are plotted in a convenient manner and then connected by a smooth curve. It is helpful to nondimensionalize the magnitude function by dividing by a reference amplitude ratio. The sensitivity, when it is the amplitude ratio at zero frequency, is a suitable reference. One method of representing the data is to plot two curves as a function of frequency, one for the nondimensional amplitude ratio, and the other for the dynamic response angle. Alternatively, the frequency measurements may be plotted on a polar diagram with the magnitude as the radius and the argument as the angle. The plotted forms of the measurements are often used without any further processing as a way to represent performance. The reason is that many techniques have been developed for interpreting and predicting the performance of systems from the frequency characteristics of systems. The techniques are based on the concepts and consequences of Fourier series and Fourier transforms. They use the performance functions in graphical rather than analytical form.

An experimental technique exists for reducing the graphical form of a performance function to analytical form. The performance function may be expressed as a ratio of two polynomials

$$[PF]_{(\text{mos})[d; c]} = \frac{P_s(jw)}{P_r(jw)} \ldots \ldots \ldots \ldots [61]$$

From Equation [26], the denominator polynomial is identified as the characteristic equation. From well-known algebraic theorems, each of the two polynomials can be represented as products of linear and quadratic factors with real coefficients, the quadratic factors having complex roots. By suitable manipulation the performance function can be brought into the form shown by the following equation from reference (13).

$$[PF]_{(\text{mos})[d; c]} = S_{(\text{mos})[d; c]}(jw)^r \prod_{i=1}^{nlf} \left[1 + \frac{jw}{(bp)_k} \right]^s$$
$$\prod_{k=1}^{nqf} \left[1 + 2\zeta_k \frac{jw}{(bp)_k} + \left(\frac{jw}{(bp)_k} \right)^2 \right]^t \ldots \ldots [62]$$

where

$S_{(\text{mos})[d; c]}$ = sensitivity of the multicoupled operating system for $q_{\text{in}(d)}$ to $q_{\text{out}(c)}$
r, s, t = integers either positive or negative
nlf = number of linear factors

nqf = number of quadratic factors

bp = break point

II means "the product of the factors"

For linear factors the break point is given by the expression

$$(bp)_i = \frac{1}{\tau_i} \dots\dots\dots\dots\dots [63]$$

where τ_i is the characteristic time. For quadratic factors the break point is the undamped angular natural frequency of the factor

$$(bp)_k = w_{n(k)} \dots\dots\dots\dots\dots [64]$$

The logarithmic form of Equation [62] is

$$\log [PF]_{(mos)[d,\ c]} = \log S_{(mos)[d,\ c]} + r \log (jw)$$
$$+ \sum_{i=1}^{nlf} s \log \left[1 + \frac{jw}{(bp)_i} \right]$$
$$+ \sum_{k=1}^{nqf} t \log \left[1 + 2\zeta_k \frac{jw}{(bp)_k} + \left(\frac{jw}{(bp)_k} \right)^2 \right] \dots\dots [65]$$

The quantities represented in Equation [65] are complex numbers. The real parts are the log magnitudes of the factors of Equation [62], while the imaginary parts are the arguments of the factors of Equation [62]. The two plots showing the variation of the log magnitude of the performance function and the argument of the performance function with the log of the frequency are often called Bode diagrams. The Bode diagrams of linear or first-order functions all have the same shape independent of the break point. Bode diagrams of quadratic or second-order functions for a given damping ratio ζ are also unique and independent of the break point. The break points determine the location of the first and second-order functions on the Bode diagrams with respect to the log frequency abscissa. Plastic profiles of the first and second-order functions like those in Fig. 10 are available, all to the same scale. When the Bode diagrams for an experimentally determined performance function are prepared, it is possible to carry out a curve-fitting process using the plastic profiles. The constructed curves are the sums of appropriately located first and second-order function Bode diagrams as indicated by Equation [65]. The break points and damping ratios are determined with the aid of suitable indexes on the plastic profiles.
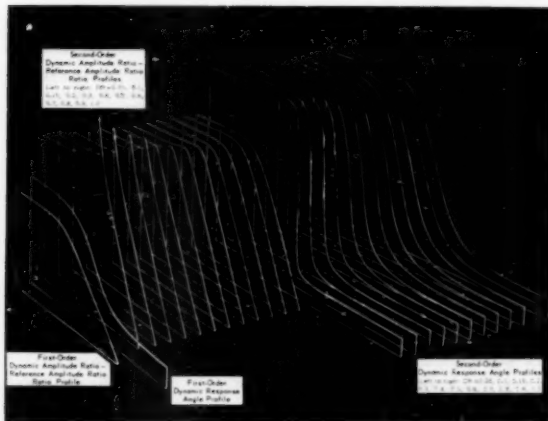
Fig. 10 First and second-order profiles, from Chapter 24, vol. 2, of Draper, McKay, and Lees (1)

Further details are given in reference (13). The plastic profiles may be used to validate the experimentally determined coefficients of Equation [51]. The break points can be calculated by following the method of the cited reference. Bode diagrams can be constructed with the profiles and compared with additionally measured values of the frequency response.

The direct measurement of steady-state frequency responses has been particularly effective in testing electrical circuits and communication systems. In other areas, experimental difficulties have limited its usefulness. Sometimes it has proved impossible to set up a steady-state sinusoidal variation either because of the physical situation or because the system does not perform well for this kind of input. Steady-state frequency measurements on guided missiles in flight are clearly impractical because of the short time available for testing. Sinusoidal variations of the inputs to chemical processing plants must have such long periods that it is difficult to maintain the equilibrium reference condition. As usual, the test data generally exhibit uncertainty, signal corruption, and distortion due to nonlinearity. The undesired effects make it particularly difficult to measure dynamic-response angles accurately. It is to be noted that when the dynamic properties of a system become important in the frequency response, the outputs become attenuated. The tendency of the person testing the system is to increase the input amplitude in order to get a readable response in the presence of the uncertainty and noise when the response amplitude gets small. However, the increased input quantity amplitude may increase the distortion due to nonlinearity, with the result that data may be of little real value. Cowley has pointed out (14) that the uncertainty makes the input signal difficult to read at very low frequencies, while the uncertainty affects the output quantity more markedly at high frequencies. He concluded that direct frequency measurements are usable only for a relatively small frequency band. Direct frequency measurements are often time-consuming. It is necessary to wait until steady-state conditions exist, which means at least two cycles. When the system is slow, the period per cycle for the low frequencies may be inordinately long.

Standard Curves Method

When the theory of the physical system is well known and the form of the performance equations is at hand, it is possible to calculate the responses for all variations of the parameters for a particular input. The dynamic measurements are determined by identifying the actual response with one of the calculated responses. The method is particularly convenient in graphical form. The plastic profiles of the first and second-order Bode diagrams referred to in the previous method typify the standard curves method.

The method is especially powerful in interpreting and evaluating the responses of first and second-order systems. Nondimensional curves for the responses of this class of systems, for a variety of inputs, are presented in "Instrument Engineering" (1). All the curves are printed to scale so that values may be read off directly. In addition, techniques and graphs are provided that allow the parameters to be determined in short order. For example, the properties of second-order systems can be evaluated by one of the methods illustrated in Fig. 11 together with the curves of Figs. 12 and 13.

The techniques developed for first and second-order systems can be applied to higher-order systems when the modes are sufficiently distinct. For example, if one mode outlasts all the others, the parameters associated with this mode are evaluated first. The system is then excited to exhibit the next slowest mode and so on.

The discussion presented here on the use of standard curves is quite brief, but the method is far more important than the short

TYPE OF RESPONSE	RESPONSE CURVE	EQUATION PARAMETERS	METHOD USED TO FIND EQUATION PARAMETERS
Oscillatory $\{ DR < 0.5 \quad \zeta < 0.5 \}$	(DVR)$_0$ can be for any peak ...	$\{ DR , T \quad \zeta , T \}$	Measure T, (DVR)$_0$, (DVR)$_1$, (DVR)$_2$, (DVR)$_3$, $\cdots$ Form ratios $\dfrac{(DVR)_1}{(DVR)_0}$, $\dfrac{(DVR)_2}{(DVR)_1}$, $\dfrac{(DVR)_3}{(DVR)_2}$, $\dfrac{(DVR)_2}{(DVR)_0}$, $\dfrac{(DVR)_3}{(DVR)_1}$, $\dfrac{(DVR)_3}{(DVR)_0}$ Find $\{ \dfrac{DR}{\zeta} \}$ from Fig. 3
Critically Aperiodic Region $\{ 0.5 < DR < 2.0 \quad 0.5 < \zeta < 2.0 \}$	Critically aperiodic curve / Typical curve	$\{ DR , \omega_n \quad \zeta , \omega_n \}$	Measure t_1, t_2, t_3 Form ratios $\dfrac{t_2}{t_1}$, $\dfrac{t_3}{t_1}$, $\dfrac{t_3 - t_2}{t_2 - t_1}$ Find $\{ DR , \omega_n t_1 , \omega_n t_2 \quad \zeta , \omega_n t_1 , \omega_n t_2 \}$ from Fig. 4 Compute value of $\{ \omega_n \}$
Critically Aperiodic $\zeta = DR = 1.0$		$\{ (CT)_{(1dr)} \quad \tau \}$	Measure $(CT)_{(1dr)} = \tau$ (For $\{ DR \zeta \} = 1.0$, $\{ (CT)_{(1dr)} \tau \} = t_1 = t_2 - t_1 = t_3 - t_2$)
Nonoscillatory $\{ DR > 1.0 \quad \zeta > 1.0 \}$	Semilog Plot of Response Curve	$\{ [(CT)R]_{(hdr)} \quad (CT)_{(hdr)(larger)} \quad \nu , \tau_1 \}$	Plot response curve on semilog paper. Extrapolate straight-line portion to establish (DVR)$_{(ext)}$ = extended dependent variable ratio. Measure log (DVR)$_1$ and log (DVR)$_2$ at t_1 and t_2 respectively. From plot take (DVR)$_0$ and (DVR)$_{(ext)}$. Compute $\{ (CT)_{(hdr)(larger)} \tau_1 \}$ from $\{ (CT)_{(hdr)(larger)} \tau_1 \} = \dfrac{t_2 - t_1}{\ln (DVR)_1 - \ln (DVR)_2}$ Note that $\ln (DVR)$ = natural logarithm of DVR $= 2.303 \log (DVR)$ Compute $\{ (CT)R_{(hdr)} \nu \}$ from $\{ (CT)R_{(hdr)} \nu \} = \dfrac{(DVR)_{(ext)}}{(DVR)_{(ext)} - (DVR)_0}$

Fig. 11 Determination of differential equation parameters and coefficients from step function response curves for the linear second-order differential equation with constant coefficients $a_2 \ddot{v} + a_1 \dot{v} + a_0 v = 0$, from Chapter 19, vol. 2, of Draper, McKay, and Lees (1)

space devoted to it might indicate. The brevity is occasioned by the simplicity of the method rather than its usefulness. The other methods require more description because they are less familiar. It is quite likely that standard curves are more widely used in interpreting dynamic measurements than any of the others. The technique of nondimensionalizing the co-ordinates or the use of normalization can make a curve more generally useful than if the co-ordinates retain dimensional values.

Indirect-Frequency Method

There are several procedures for obtaining the harmonic content of an arbitrary variation of a quantity with respect to time. These procedures permit the frequency performance functions to be produced from pulse inputs, step inputs, or even random inputs. In other words, the frequency response characteristics of a system may be obtained by indirect means rather than by sinusoidal forcing. It is possible to circumvent many of the limitations of direct frequency measurements such as the short life of guided missile systems.

The mathematical basis relating the time variation to the frequency function is the Fourier transform. The relationship may be written as

$$[FT]q(t) = q(w) = \int_{-\infty}^{\infty} q(t)e^{-jwt}dt \dots \dots [66]$$

When the quantity is zero for negative values of time and returns to zero in a finite interval T the Fourier transform may be calculated from an approximation

$$[FT]q(t) \approx \sum_i q(i\Delta t)e^{-jw(i\Delta t)}(CF) \dots \dots [67]$$
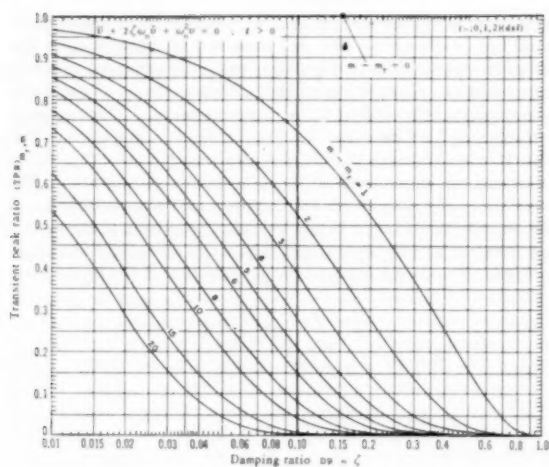
a) Logarithmic scale abscissa – linear scale ordinate plot

b) Logarithmic scale plot

Fig. 12 Transient peak ratio as a function of damping ratio for the
$(-1; 0, 1, 2)$ linear second-order differential equation with constant
coefficients when the step-function solution is oscillatory (from
Chapter 19, vol. 2 of Draper, McKay, and Lees (1)
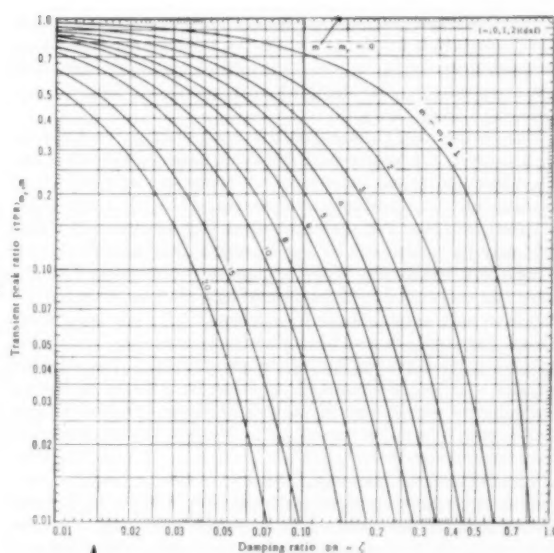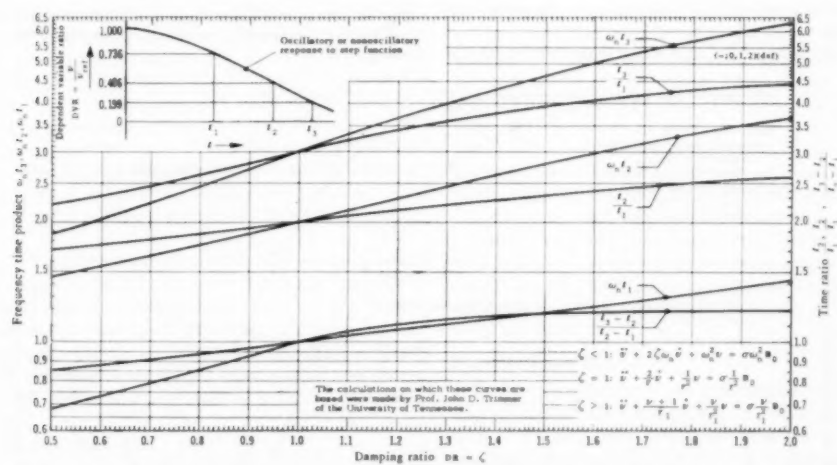
Fig. 13 Chart for determining the damping ratio and the undamped angular natural frequency from the
step-function response for the $(-1; 0, 1, 2)$ linear second-order differential equation with constant co-
efficients when the damping ratio is between 0.5 and 2.0, from Chapter 19, vol. 2, of Draper, McKay,
and Lees (1)

where

$q(i\Delta t)$ = a representative value of the quantity in the interval
$i\Delta t$ to $(i + 1)\Delta t$

(CF) = correction function depending on the method of ap-
proximating the time history of the quantity

The proper mathematical conditions are satisfied when the input
quantity is a pulse and the output returns to its reference value.
The Fourier transform can be calculated for both the input and

output quantities in such a way that the same correction function
appears in each calculation. The performance function may be
obtained from the expression

$$[PF]_{(\text{mos})[\text{d el}]} = \frac{\sum_{i} q_{\text{out}(p)}(i\Delta t)e^{-jw(i\Delta t)}}{\sum_{k} q_{\text{out}(q)}(k\Delta t)e^{-jw(k\Delta t)}} \quad \cdots \cdots [68]$$

Since both the numerator and denominator are sums of complex

numbers, the real parts and the imaginary parts must be summed separately. It is helpful to reduce the performance function to its magnitude and argument. When the calculations are carried out, the results may be identified with the relationships of Equations [53], [54], and [55].

The results may be interpreted by any of the procedures described under the steady-state frequency measurements method. The discussion at this point is concerned mostly with techniques for calculating the performance function.

Direct calculation of the summations that appear in Equation [68] may be carried out by hand using a desk calculator or by means of a large-scale digital computer. A large-scale digital computer is almost mandatory when there are a hundred or more points in each summation. In addition to the general-purpose digital computer, special-purpose analog computers have been developed for carrying out the work (15, 16). The 24-point Fourier synthesizer described in reference (15) is of limited capability because it can only handle 24 terms. Another Fourier transformer is being perfected, which has been demonstrated to have considerable potentiality (17).

The distinction between a Fourier series and a Fourier transform for a function becomes somewhat blurred when the duration of the function is finite. This is particularly true when the Fourier transform is calculated numerically for a finite number of frequencies. The literature is replete with methods and descriptions of Fourier series computers. As an example, a Swiss-made mechanical device, the Coradi analyzer, calculates Fourier series coefficients on the principle of a planimeter with an ultimate precision of at least 0.2 per cent of the average value of the signal. Another Fourier analyzer that is reported to perform with a precision comparable to the Coradi analyzer is described by F. J. McDonal (18). The Fourier analyzers produce both magnitude and phase information as needed to carry out the coefficient evaluation program.

An illustration of the capabilities of Fourier transform calcula-

tions is given in Figs. 14 and 15. The system being tested was a small heat exchanger. The test consisted of introducing a disturbance into the feed line and measuring the resultant temperature variation. Fig. 14 shows the records of the input valve displacement and the effluent temperature variation. The frequency performance function corresponding to the data is shown in Fig. 15. It was fortunate that it was also possible to obtain direct frequency measurements of the process for comparison. The direct frequency measurement data are also shown in Fig. 15. The drawn curves are obtained by the curve-fitting process using the plastic profiles mentioned under steady-state frequency method. The curves correspond to a function of the form

$$S_{(\Delta\omega)} \frac{1 - 2\zeta \dfrac{j\omega_f}{\omega_n} + \left(\dfrac{j\omega_f}{\omega_n}\right)^2}{(1 + \tau_1 j\omega_f)(1 + \tau_2 j\omega_f)} e^{-j(dt)\omega_f} \dots\dots [69]$$

where

$\omega_n = 1.25$ radians/sec $dt = 1.75$ sec
$\tau_1 = 7.6$ sec $\zeta = 0.3$
$\tau_2 = 1.41$ sec
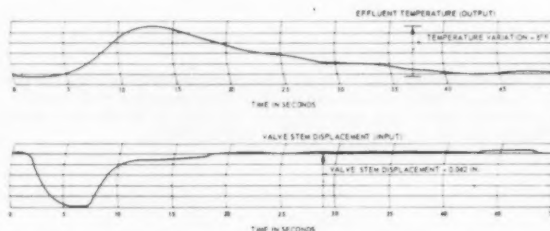
Fig. 14   Record of input pulse to a model heat exchanger and resulting output temperature variation, from Fig. 5 of Lees and Hougen (24)
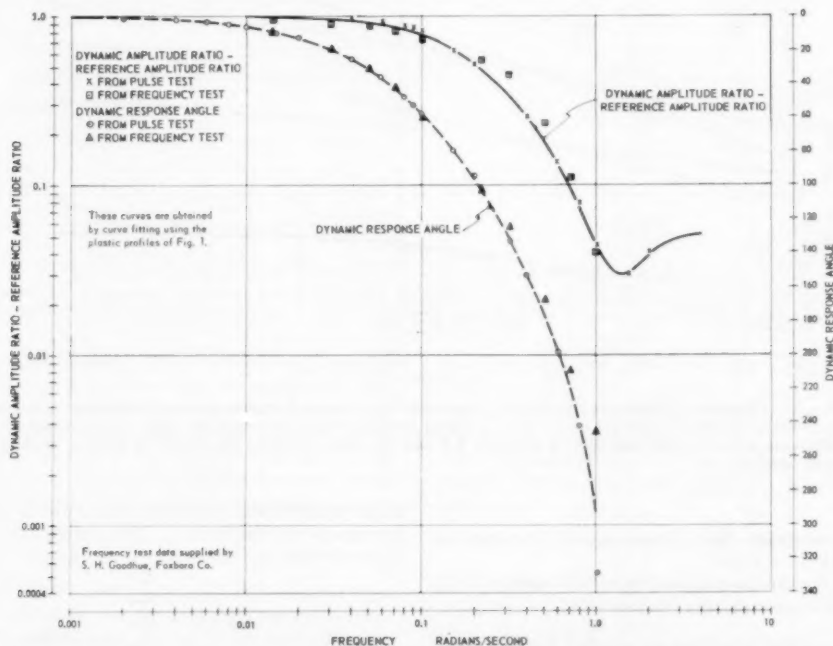
Fig. 15   Frequency performance function for model heat-exchanger process, from Fig. 6 of Lees and Hougen (24)

The form of Equation [69], and the numerical values, were found by curve fitting.

In some applications, as in vibration analysis and in some communication system requirements, it is sufficient to determine the amplitude distribution of the time variations of the output quantities. For these applications there exist special purpose spectrum analyzers. One class of spectrum analyzers uses sharply tuned filters. When the quantity being analyzed is a mechanical vibration, the sharply tuned filters may be a group of reeds firmly mounted to a common base. The amplitude of the forcing-component frequency is determined by the amplitude of the sensing-reed oscillation. Electrical quantities may be analyzed by converting to mechanical motion through a suitable transducer. Modern refinements have changed the reeds to other mechanical filters, such as crystals and rods. In addition, piezoelectric and magnetostrictive filters have both electrical inputs and outputs. Other designs employ electrical circuit filters like inductance-capacitance circuits or twin-T resistance-capacitance circuits. The Q of a filter is a measure of the band-pass characteristics of the filter, in the sense that the greater the value for Q the narrower is the range of frequencies the filter will admit to pass. The Q of filters designed for passing higher frequencies can be made much greater than for filters passing lower frequencies. When the quantity being analyzed has only low-frequency components it is helpful to employ a technique that will change the frequency range of the quantity without altering the amplitude relationships among its harmonic components. In one scheme the signal is heterodyned with a high-frequency reference signal. In another scheme, where the quantity is available in recorded form, particularly when it is recorded on magnetic tape, the frequency components may be shifted by speeding up the rate of scanning the record.

A commonly used design employs only one sharply tuned filter. The reference heterodyning frequency is varied over a band that includes all the desired components. The heterodyned signal is passed through the filter to measure the component equal to the difference between the reference frequency and the filter peak frequency. An alternative design is described by Duncan (19). In this analyzer 420 magnetostrictive filters are employed simultaneously. The heterodyned signal is fed into all the filters at once and the filters are sampled by a high-speed switch. The output is presented as a continuous cathode-ray-oscilloscope pattern.

Another class of spectrum analyzers uses an electrical circuit that effectively computes the Fourier series component amplitude (20). Demodulators are circuit components that have an output equal to the product of a reference electrical signal of constant frequency and amplitude and an arbitrary electrical signal. When the demodulator output is averaged by conventional resistance-capacitance circuits, the result is proportional to the Fourier series coefficient

$$e_{out} = \frac{S_{dem}}{T} \int_0^T e_{in}(t) \sin w_r t \, dt \ldots \ldots \ldots [70]$$

where

$e_{in}(t)$ = arbitrary input signal
w_r = reference angular frequency
T = period of integration
S_{dem} = demodulator sensitivity

A heterodyne technique may be employed to improve the performance characteristics of the demodulator-integrator circuit. The arbitrary input signal is surveyed by sweeping the heterodyne reference frequency. The output of the demodulator circuit may be presented as a continuous cathode-ray-oscilloscope pattern. When it is convenient, the scope picture may use log-log presentation by the use of appropriate electrical networks.

The indirect frequency methods can reduce the effects of signal corruption but not necessarily the uncertainty of the measured quantity. The spectrum analyzers using narrow band-pass filters pass only the noise within their narrow frequency range. Since the noise power in a narrow range is small, the signal-to-noise ratio may be improved when the signal has a strong component in the test frequency range. The techniques employing Fourier transforms, Fourier series, or demodulator circuits carry out an averaging process with respect to a single frequency or possibly a very narrow frequency range, that tends to reduce the effect of noise. In addition, these techniques measure only the properties of the quantity for the test frequency component. The effect of nonlinear distortion may be interpreted as introducing higher frequency (as well as subfrequency) components. The indirect frequency methods are generally sufficiently selective to reject the nonlinear distortion components. There is a question, of course, whether the system performance is properly described by the theory when distortion exists. The question can only be answered by close examination of the behavior of the system under study.

A further benefit of the indirect frequency methods is the potential economy of system time effected by the use of pulse inputs. The pulse technique as an indirect frequency method has been successfully applied to aircraft, fire-control systems, and heat exchangers.[5] In the original applications rectangular-shaped pulses were thought to be best suited as the test function, but actually little consideration was given to the problem. Later, Smith and Triplett discussed the influence of pulse shape with respect to the frequency content. In addition to the frequency considerations raised by Smith and Triplett, when the input changes too violently there exists the potentiality of saturation that was mentioned in the previous paper. The selection of a suitable pulse depends on choosing an effective pulse shape, with the proper duration, and with adequate strength. The pulse shape is chosen to maximize the pulse harmonic content in the response region of the system being tested. The effect of the pulse shape on its harmonic content is shown in Fig. 16. The pulse duration must be short enough to excite the system dynamic properties but not so short as to cause it to saturate. Preliminary experimental tests help establish these limits. The pulse strength is the area of the pulse and is made large enough to minimize uncertainties, without saturating the system. For systems with a dominant second-order mode, Coppedge and Wolf (25) have shown that a pulse having a continuous first derivative should have a duration about equal to the period of the second-order mode. The practical limits of the pulse method are still being studied.

Random Input Method

The previous methods for determining system parameters and characteristics have all required the use of particular input functions. These input functions must be generated by signal generators and injected into the appropriate input receiver of the system. An indirect method has been proposed to use the records of the inputs and outputs obtained in the normal operation of the system. The method requires the presence of random variations in the input quantities. In many instances the naturally occurring random variations are strong enough to excite the dynamic characteristics of the system. In those applications where the naturally occurring random variations are inadequate, it has been proposed to superimpose random variation on the normal operating signal. "White noise" signal generators have been developed for this purpose for communication-system applications.

The formal basis for random input techniques lies in the convolution integral expression relating the cth output to the dth in-

---

[5] See References (21, 22, 23, and 24).
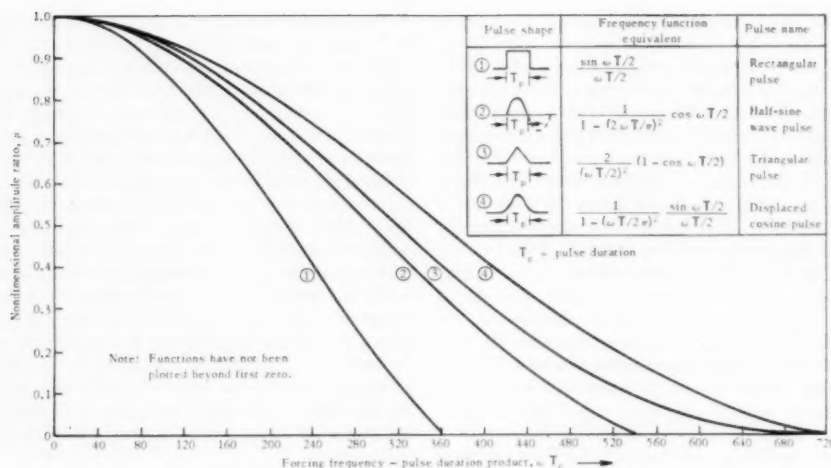
Fig. 16 Comparison of harmonic content of typical pulse shapes

put of a multicoupled operating system as mentioned in the previous paper

$$q_{out(c)} = \int_{-\infty}^{t} [WF]_{(mos)[d;\,c]}(t - t_c)q_{in(d)}(t_c)dt_c \ldots [71]$$

where

$[WF]_{(mos)[d;\,c]} = d$th input to cth output weighting function for the multicoupled operating system

The task is to determine the weighting function from the records of the dth input and cth output. For the situation where the ideal values of the input and output are known, unencumbered by uncertainties and signal corruptions, it is possible, at least in theory, to extract the weighting function from the relationships of Equation [71]. A mathematical technique was proposed by Travers (26) in 1949, using correlation functions.

The autocorrelation function of the input is defined as

$$[ACF]_{q_{in}}(x) = \lim_{T \to \infty} \frac{1}{2T}\int_{-T}^{T} q_{in}(t)q_{out}(t + x)dt \ldots [72]^6$$

The cross-correlation function between the input and output is defined as

$$[CCF]_{[q_{in};\,q_{out}]}(x) = \lim_{T \to \infty} \frac{1}{2T}\int_{-T}^{T} q_{in}(t)q_{out}(t + x)dt \ldots [73]^6$$

The expression in Equation [71] may be rewritten by a change of the integration variable as

$$q_{out} = \int_{0}^{\infty} [WF](t_c)q_{in}(t - t_c)dt_c \ldots [74]$$

Subscripts indicating the particular input and output are dropped because the discussion at this point is unambiguous. By introducing Equation [74] into Equation [73] and interchanging the order of the operations, Travers has shown that

$$[CCF]_{[q_{in};\,q_{out}]}(x) = \int_{0}^{\infty} [WF](t_c)[ACF]_{q_{in}}(x - t_c)dt_c \ldots [75]$$

The range of the variable x extends to both negative and posi-

6 The symbols commonly used for the correlation function are
ϕ_{ii} for the input autocorrelation function
ϕ_{io} for the input-output cross-correlation function
However, the symbol ϕ has already been used twice in this paper—once for the phase in transient response components and again for the dynamic response angle. The notation adopted here avoids confusion between correlation functions and phase angles.

tive values. The form of the expression in Equation [75] is identical to that in Equation [74], but relates functions derived from the input and output quantities instead of these quantities themselves.

It is to be observed that the relationship in Equation [75] employs the correlation functions that are usually associated with statistical processes. However, Equation [75] has not been developed from a statistical point of view and is not to be so interpreted. It is rather a mathematical technique for extracting the weighting function. A typical random input record and the response are shown in Fig. 17(a), where the input is labelled "error" and the output is "handle velocity." A typical autocorrelation function [not derived from the record of Fig. 17(a)] and a typical cross-correlation function are shown in Figs. 17(b) and (c). The mathematical procedure is seen to convert the input into a pulse-like function and the output into the response to a pulse-like function. The advantage of the random input technique based on Equation [75] rests on the relative ease in treating pulse functions as compared with the original input.

One method for the direct extraction of the weighting function employs an approximation for the integral in Equation [75]

$$[CCF]_{[q_{in};\,q_{out}]} = \sum_{i=0}^{\infty} [WF](i\Delta t_c)(x - i\Delta t_c)\Delta t_c \ldots [76]$$

Tustin (27) invented the concept of the time series for representing the terms that appear on the right-hand side of Equation [76] and for calculating the convolution integral with time series. Madwed (28) developed a theoretical basis for time series techniques in terms of a number series transformation. In Madwed's notation, the convolution integral is

$$[NST][CCF]_{[q_{in};\,q_{out}]} = [NST][WF] \times [NST][ACF]_{q_{in}} \ldots [77]$$

By cross-division, Madwed means the process of producing the time series representing the weighting function. In his notation

$$[NST][WF] = \frac{[NST][CCF]_{[q_{in};\,q_{out}]}}{[NST][ACF]_{q_{in}}} \ldots [78]$$

In the cited reference Madwed presents a straightforward method by carrying out the cross-division. The process can be done either by hand or by digital computers.

Y. W. Lee has pointed out that the autocorrelation of truly white noise is an impulse function
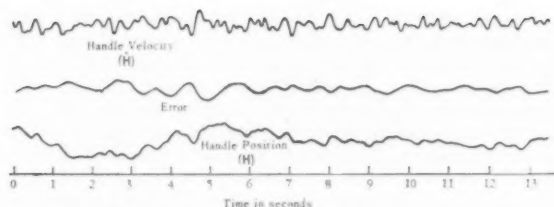
Fig. 17(a)   Tracking record obtained at University of Birmingham, based on Fig. 11 of Goodman and Reswick (34)
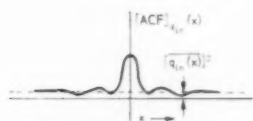
Fig. 17(b) Typical autocorrelation function, based on Fig. 5(d) of Goodman and Reswick (34)
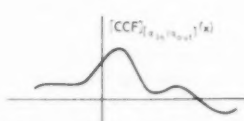
Fig. 17(c) A typical cross-correlation function, based on Fig. 6 of Goodman and Reswick (34)

$$[ACF](\text{white noise}) = [IF] \dots \dots \dots \dots [79]$$

For those applications where white noise can be injected as the input, the weighting function is found directly from the record, and the integral of Equation [75] is

$$[WF] = \frac{[CCF]_{[q_{in};\ q_{out}](\text{white noise})}}{[IFS]} \dots \dots \dots [80]$$

where

$$[IFS] = \text{impulse function strength} = \int_0^\infty [ACF](\text{white noise})dx$$

$[CCF]_{[q_{in};\ q_{out}](\text{white noise})} = $ cross correlation between the white noise input and the output due solely to this input

Travers (26) has suggested that the weighting function be extracted by calculating the Fourier transform of Equation [75]. From theory, the transform of a convolution integral is a product

$$[CCF]_{[q_{in};\ q_{out}]}(j\omega) = [WF](j\omega)[ACF]_{q_{in}}(j\omega) \dots [81]$$

$$[WF](j\omega) = [PF](j\omega) = \frac{[CCF]_{[q_{in};\ q_{out}]}(j\omega)}{[ACF]_{q_{in}}(j\omega)} \dots [82]$$

It is noted that in Equation [82] the frequency form of the weighting function is the frequency form of the performance function. When the random input method is used to produce the Fourier transforms of the correlation functions, it becomes another indirect frequency method. In fact, it can be shown that

$$[ACF]_{q_{in}}(j\omega) = \frac{\pi}{T}\ q_{in}(-j\omega)q_{in}(j\omega) \dots \dots [83]$$

$$[CCF]_{[q_{in};\ q_{out}]}(j\omega) = \frac{\pi}{T}\ q_{in}(-j\omega)q_{out}(j\omega) \dots [84]$$

so that, from Equation [82]

$$[PF](j\omega) = \frac{q_{out}(j\omega)}{q_{in}(j\omega)} \dots \dots \dots \dots [85]$$

A straightforward procedure is to calculate the correlation functions and then calculate the corresponding Fourier transforms. The performance function is obtained from Equation

[82]. Elkind (29) has developed a cross-spectrum computer that obtains the Fourier transforms directly using Equations [83] and [84].

These procedures yield the weighting function as a time series or as a frequency function. When the system parameters are desired, it is necessary to perform additional computations. The frequency function can be reduced to analytical form by one of the procedures described previously. The weighting function is the response to a unit impulse. For a linear system it is a sum of exponentials with exponents having the roots of the characteristic equation as factors. Shinbrot (5) suggests the case of Prony's method and he includes a description of the technique in his report. The performance equation coefficients may be evaluated by the procedure included in the measurement of the derivatives method.

It has been postulated that the records of the input and output quantities are pure and free from uncertainty and signal corruptions. Such a situation cannot really occur, and it is necessary to take into account the influence of the limitations of measurement as has been done for the other methods. The random input variations are distinguished in theory from the uncertainties, and it is assumed that the distinction is sufficient for a separation to be effected. One of the benefits claimed for the random input method is that the technique permits the incorporation of statistical procedures to isolate the desired quantities. In order to illustrate the statistical aspect, let it be assumed that the measured input and output quantities are representable, respectively, as the sum of the desired quantity plus deviations

$$q_{in(m)} = q_{in} + (D)q_{in} \dots \dots \dots [86]$$

$$q_{out(m)} = q_{out} + (D)q_{out} \dots \dots \dots [87]$$

where the subscript m indicates the measured quantity. By "deviation" is meant the contributions from all sources not due to the actual quantity.

The correlation functions for the measured quantities are

$$[ACF]_{q_{in(m)}} = [ACF]_{q_{in}} + [ACF]_{(D)q_{in}} + [CCF]_{[q_{in};\ (D)q_{in}]} + [CCF]_{[(D)q_{in};\ q_{in}]} \dots \dots [88]$$

$$[CCF]_{[q_{in(m)};\ q_{out(m)}]} = [CCF]_{[q_{in};\ q_{out}]} + [CCF]_{[q_{in};\ (D)q_{out}]}$$

$$+ [CCF]_{[(D)q_{in},\ q_{out}]} + [CCF]_{[(D)q_{in};\ (D)q_{out}]} \dots \dots [89]$$

The relationship between the cross-correlation function and the autocorrelation function developed as Equation [75] is not generally valid for the correlation functions of the measured quantities expressed in Equations [88] and [89]. It is necessary to be able to evaluate and isolate the contributing correlation functions due to the deviations. It is often assumed that the deviations are statistically incoherent with the desired quantities and also with each other. This assumption is equivalent to assuming that the cross-correlation functions of the deviations must be zero. Accordingly, Equations [88] and [89] become

$$[ACF]_{q_{in(m)}} = [ACF]_{q_{in}} + [ACF]_{(D)q_{in}} \dots \dots [90]$$

$$[CCF]_{[q_{in(m)};\ q_{out(m)}]} = [CCF]_{[q_{in};\ q_{out}]} \dots \dots \dots [91]$$

It is still necessary to establish a basis for the removal of the autocorrelation function of the input quantity deviations before the correlation functions of the measured quantities may be introduced into Equation [75]. It may be assumed from a practical point of view that careful control of the measuring and recording processes can reduce the input deviations to negligible value. It is still necessary to be assured that the autocorrelation function is negligible within the required precision.

The assumptions of statistical incoherence and negligibility present several difficulties. The correlation functions presume records of infinite length, or at least of sufficient duration, to be statistically valid. It is necessary to have a complete record of the statistical variation of the deviations as well as the random variations that are the true input and output quantities. However, a long record requires a lengthy calculation to obtain the correlation functions, and it is tempting to shorten the record in the interest of economy. The exact relationship between record length and the reduction of the effects of the deviations is not very well known. Margolis (30) has indicated that the reduction in the influence of the deviations is proportional to the inverse square root of the record duration.

An alternative approach to the random input method that avoids the need for the assumptions about the nature of the deviations may be developed from the least-squares concept. The error between the calculated and measured outputs is defined as

$$(E)q_\text{out} = q_\text{out} - q_{\text{out}(m)} \dots \dots \dots \dots [92]$$

The best information concerning the input quantity is the measured input including the deviations. The relationship between the calculated output and the input is

$$q_\text{out} = \int_0^\infty [WF](t_c)q_{\text{in}(m)}(t - t_c)dt_c \dots \dots [93]$$

From Equations [92] and [93]

$$(E)q_\text{out} = \left\{ \int_0^\infty [WF](t_c)q_{\text{in}(m)}(t - t_c)dt_c \right\} - q_{\text{out}(m)} \dots [94]$$

The procedure is to seek the minimum value of the mean square error because this is the only practical criterion available at this time. The definition of the mean square error is

$$\bar{E} = (MSE) = \frac{1}{2T}[(E)q_\text{out}]^2$$

$$= \frac{1}{2T}\int_{-T}^T [q_\text{out} - q_{\text{out}(m)}]^2 dt \dots \dots [95]$$

where T is the duration of the record. It is not assumed that the record must be of a duration comparable to infinity for the system.

The mean square error is identical in concept with the error square conditions given by Equation [42]. The difference in the forms is that the measured and computed functions in Equation [42] refer to a controlled variation of the input like a step function, while the functions in Equation [95] represent the response to a randomly varying input.

It has been shown that the minimum mean square error is obtained when the Wiener-Hopf equation is satisfied (31, 32, 33)

$$[CCF]_{(q_\text{in};\ q_\text{out})}(x) = \int_0^\infty [WF](t_c)[ACF]_{q_\text{in}}(x - t_c)dt_c \dots [96]$$

$$\text{for } x > 0$$

The Weiner-Hopf equation[7] is almost identical to the relationship Travers developed shown in Equation [75]. It differs, however, in being valid only for positive values of the running variable x and in that the correlation functions are computed only for finite duration. A typical autocorrelation function is shown in Fig. 17(b). The integral in Equation [96] is calculated for all values of the running variable, including negative ones. The cross-correlation function need only match the integral for positive values of the running variable. When the weighting function found from Equation [96] remains the same for records of various durations, it may be assumed that the weighting function is applicable.

[7] See Appendix for the derivation of the Wiener-Hopf equation.

The technique devised by Shinbrot for determining the weighting function by the least-square condition described under the derivatives method can be used to extract the weighting function from Equation [96]. Margolis (30) has developed a method based on the number series approximation given by Equation [76]. He presents the expression as a series of linear simultaneous equations by writing a separate equation for each value of the running variable x. The unknowns are the values of the weighting function for values of $(i\Delta t_c)$. The number of linear simultaneous equations is equal to the number of terms in the series required to represent the weighting function. More detail requires a larger number of equations. The terms in the weighting function time series are found by any of the algebraic processes for solving linear simultaneous equations.

Goodman and Reswick (34) describe an experimental procedure for solving Equation [96] using the weighting function time series formulation. They use a delay line synthesizer previously developed by Reswick. The delay line synthesizer is an analog-type computer especially adapted to use the weighting function expressed as a time series. This type of analog computer was first conceived by J. M. L. Janssen and L. Ensing (35).

Goodman and Reswick generate the correlation functions by an analog technique, with the aid of a photoelectric dual-function generator of their own design, from the input and output quantity records. The correlation functions are then set into the dual-function generator. The delay line synthesizer controls are adjusted to bring about the condition required by Equation [96], using a cathode-ray oscilloscope to indicate a null condition between the computed and measured cross-correlation functions for positive values of the running variable. This delay line synthesizer has the capacity of twenty terms in the time series.

The discussion of the random-input method so far has dealt only with a system having a single input and a single output. For the multicoupled operating system with several inputs and outputs it is shown in the previous paper that the cth output is given by the summation

$$q_{\text{out}(c)} = \sum_{d=1}^r \int_0^\infty [WF]_{(\text{mos})[d;\ c]}(t_c)q_{\text{in}(d)}(t - t_c)dt_c \dots [97]$$

For the ideal situations when there is neither uncertainty nor signal corruption it is possible to develop a set of relationships comparable to Equation [75]

$$[CCF]_{[q_{\text{in}(e)};\ q_{\text{out}(c)}]} = \sum_{d=1}^r \int_0^\infty [WF]_{(\text{mos})[d;\ c]}$$

$$(t_c)[CCF]_{[q_{\text{in}(e)};\ q_{\text{in}(d)}]}(x - t_c)dt_c$$

$$e = 1, 2, \dots, r \dots \dots \dots \dots \dots [98]$$

When the eth input is identical to one of the other inputs, the cross-correlation function is an autocorrelation function. In each of the equations in the set of Equations [98] there will be one autocorrelation function. It is possible, in theory, to extract all the weighting functions from the equation set. For the perfect situation where all the inputs are mutually incoherent statistically, the cross-correlation functions are zero and the equation set becomes

$$[CCF]_{[q_{\text{in}(d)};\ q_{\text{out}(c)}]} = \int_0^\infty [WF]_{(\text{mos})[d;\ c]}$$

$$(t_c)[ACF]_{q_{\text{in}(d)}}(x - t_c)dt_c$$

$$d = 1, 2, \dots, r \dots \dots \dots \dots \dots [99]$$

Goodman (36) has investigated the problems of treating a multi-

input system. He has described a procedure using the delay line synthesizer for a system having two or three inputs, when the inputs are not mutually statistically incoherent. He does not treat the problem of the uncertainties and signal corruptions.

The application of the least-squares method to the multi-coupled operating system has not been carried out. It is necessary to find the appropriate extension of the Wiener-Hopf equation that may be employed in this situation.

Another problem that has not been resolved for the random-input method is the linearization of multicoupled operating systems. It was pointed out in the first paper of this series that linearization methods may be applied to systems that are non-linear in their over-all operation. Goodman in the paper cited has also attempted to treat this problem with modest success. The difficulty to be resolved is that while the linear system concept can be usefully applied to the particular situation and the results can predict the performance of the system, it is insufficient to produce a formalistic solution to represent a severely restricted set of operating conditions. It is always possible to obtain a set of numbers for one special case but inapplicable for any variation of the operating conditions.

Combination Method

The description of the techniques presented in the previous section is directed toward quantities that are best treated statistically. The techniques have greater generality, however, and may be employed equally well on either controlled or random inputs. Correlation functions may be used to reduce the effects of uncertainties and noise associated with controlled inputs in the same manner as for random inputs. The assumption about incoherence between the desired quantity and the deviations holds with more force than for random inputs.

In order to take fullest advantage of the correlation techniques, the record should be of considerable duration and the controlled input should be repeated a number of times. Y. W. Lee (37) has demonstrated that the autocorrelation function of a periodic function is also periodic. In particular, the autocorrelation function of a sinusoidal variation plus deviations is sinusoidal for large enough values of the delay. The cited reference shows that it is possible to reduce the effect of noise by a significant factor.

The cross-correlation function of a sinusoidal variation with another sinusoidal variation of the same frequency is also periodic. When the first sinusoidal function is corrupted with noise and uncertainties, the cross-correlation function of the second sinusoidal function with the noise is zero. The cross correlation of a measured sinusoidal function with a pure sinusoidal function of the same frequency should be free of corruptions even for zero delay. In the cited paper by Lee, Cheatham, and Weisner (37), it is shown that the noise was reduced by a factor of a hundred in an experimental test of the procedure for a frequency of 8 kilocycles. The technique of using a detector circuit that is mentioned under indirect-frequency methods may .be regarded as a technique for obtaining the cross correlation between two periodic functions of the same frequency with zero delay. This shows that the Fourier-series coefficient is a cross-correlation function with zero delay. Cowley (14) has developed equipment for extending the technique employed in detector circuits to frequencies as low as 0.1 cpm. He claims to have reduced the effect of noise by a factor of 250 and there is no indication that it cannot be further reduced.

As noted in the midst of the discussion of correlation functions, more elementary techniques exist for treating uncertainties and noise. For example, many applications of correlation functions require the average value of the uncertainty be zero. Averaging processes are relatively simple to instrument. Beard and Skomal (38) describe an instrument for processing fractional microvolt signals that are treated in geophysical sur-

veys. The signals are complex waveforms with repetition rates between 1 and 10 times per sec. In their scheme, the signal is sampled 72 times per cycle by a commutator switch. Each contact is coupled to a single resistor-condenser averaging circuit. Effectively, the device divides the repetition period into a suitable number of intervals, integrates the noise in each interval, while measuring the average value of the signal in the interval and stores the measured values on the capacitors. The averaged signals are taken from the capacitors and displayed on a cathode-ray oscilloscope. The authors claim a reduction of noise by a factor of 85, which is comparable to the other cited techniques.

The use of procedures combining the techniques of several methods broadens the entire field of interpreting dynamic measurements. It is difficult at this point to do more than report some of the progress.

Conclusions

The objectives and the basic methods of interpreting dynamic measurements to determine system parameters have been presented. It is believed that new techniques that will appear may be classified by one of the six methods described here. The equipment described in connection with one or the other of the methods is to be regarded as only representative. No attempt has been made to provide an exhaustive survey of apparatus.

The next task is to evaluate the methods in the light of the criteria given in the previous paper. It appears possible and feasible to make the evaluation in terms of suitable expressions for the experimental limitations.

Bibliography

1 "Instrument Engineering," by C. S. Draper, Walter McKay, and Sidney Lees, McGraw-Hill Book Company, Inc., New York, N. Y., vol. 1, 1952; vol. 2, 1953; vol. 3, part 1, 1955; vol. 3, part 2, being written.
2 "Advanced Calculus for Engineers," by F. B. Hildebrand, Prentice-Hall, Inc., New York, N. Y., 1949.
3 "Irreversibility and Generalized Noise," by H. B. Callen and T. A. Welton, *Physical Review*, vol. 83, 1941, pp. 34–40.
4 "A Survey of Methods for Determining Stability Parameters of an Airplane From Dynamic Flight Measurements," by Harry Greenberg, NACA TN 2340, 1951.
5 "A Least-Squares Curve Fitting Method With Applications to the Calculation of Stability Coefficients From Transient-Response Data," by Marvin Shinbrot, NACA TN 2341, 1951.
6 "Theory of Servomechanisms," edited by H. M. James, N. B. Nichols, and R. S. Phillips, Radiation Laboratory Series, 25, McGraw-Hill Book Company, Inc., New York, N. Y., 1949, Chapter 7.
7 "The Graphical Solution of Stability Problems," by R. K. Mueller, *Journal of Aeronautical Sciences*, vol. 4, June, 1937, p. 324.
8 "The Time Vector Method for Stability Investigations," by K. H. Doetsch, R. A. E. Report, Aero 2495, 1953.
9 "A Vector Method Approach to the Analysis of the Dynamic Lateral Stability of Aircraft," by L. Sternfield, *Journal of Aeronautical Sciences*, vol. 21, April, 1954, p. 251.
10 "Résumé of the Time Vector Method and Means of Analyzing Aircraft Stability Problems," by W. O. Breuhaus, WADC Tech. Report 52–299, November, 1952.
11 There are many references describing the uses of frequency measurements. A recent publication is "Frequency Response," edited by Rufus Oldenburger, The Macmillan Company, New York, N. Y., 1956.
12 "Response of Physical Systems," by J. D. Trimmer, John Wiley & Sons, Inc., New York, N. Y., 1950, Chapter 5.
13 Reference (1) Chapter 24.
14 "The Application of an Analog Computer to the Measurement of Process Dynamics," by P. E. A. Cowley, TRANS. ASME, vol. 79, 1957, pp. 823–832.
15 "Graphical Application of Fourier Series to Automatic Tracking Systems," by B. G. Bromberg, Report 6445-T-7, Instrumentation Laboratory, Massachusetts Intitute of Technology, Cambridge, Mass., 1948 (confidential).
16 "Torsional Fourier Transformer," by Sidney Lees, Application date June 22, 1955, Patent Application Serial No. 517314.
17 "A Torsional Fourier Transformer," by J. F. Wester, Report

T-67, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., May, 1954.

18   "A Fourier Analyzer," by F. J. McDonal, Review of Scientific Instruments, vol. 24, pp. 272–276.

19   "High-Speed, High-Resolution Spectrum Analyzer," by N. L. Duncan, Raytheon Manufacturing Company, Missile and Radar Division, Bedford, Mass.

20   A typical device in this class is the Panoramic Sonic Analyzer manufactured by Panoramic Sonic Products, Inc., Mt. Vernon, New York.

21   "The Pulse Method for the Determination of Aircraft Dynamic Performance," by R. C. Seamans, Jr., B. P. Blasingame, and G. C. Clementson, Journal of the Aeronautical Sciences, vol. 17, no. 1, January, 1950.

22   "Experimental Flight Methods for Evaluating Frequency-Response Characteristics of Aircraft," by G. A. Smith and W. C. Triplett, TRANS. ASME, vol. 76, 1954, pp. 1383–1393.

23   "Comparison of Sinusoidal Forcing Function and Pulse Forcing Function Techniques for Obtaining Dynamic Performance Characteristics of Fire Control Systems," by C. W. Ross and M. T. Gardner, Report T-30, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., May, 1953 (Confidential).

24   "Pulse Testing a Model Heat Exchange Process," by Sidney Lees and J. O. Hougen, Industrial and Engineering Chemistry, vol. 48, 1956, pp. 1064–1068.

25   "Pulse Testing of Systems With Some Nonlinearities," by T. N. Coppedge, Jr., and R. L. Wolf, Report T-68, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., May, 1954.

26   By Paul Travers, Trans. AIEE, vol. 68, 1949, p. 345.

27   "A Method of Analyzing the Behavior of Linear Systems in Terms of Time Series," by A. Tustin, Journal of the Institute of Electrical Engineers, vol. 94, Part II-A, No. 1.

28   A. Madwed, whose treatment is summarized in Chapter 27 of "Instrument Engineering" (Ref. 1).

29   "A Cross-Spectrum Computer for System Analysis," by J. I. Elkind, Group Report 38-13, Lincoln Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., April, 1955.

30   "Application of Statistical Methods to the Measurement of Industrial Process Behavior," by S. G. Margolis, SM thesis, Massachusetts Institute of Technology, Cambridge, Mass., 1955.

31   "Extrapolation, Interpolation, and Smoothing of Stationary Time Series," by N. Wiener, John Wiley & Sons, Inc., New York, N. Y., 1949.

32   "Automatic Feedback Control Systems," by J. G. Truxal, McGraw-Hill Book Company, Inc., New York, N. Y., 1955.

33   "Engineering Cybernetics," by H. S. Tsien, McGraw-Hill Book Company, Inc., New York, N. Y., 1954.

34   "Determination of System Characteristics From Normal Operating Records," by T. P. Goodman and J. E. Reswick, TRANS. ASME, vol. 78, 1956, p. 259.

35   "The Electro-Analog," by J. M. L. Janssen and L. Ensing, Philips Technical Review, vol. 12, 1951.

36   "Determination of the Characteristics of Multi-Input and Nonlinear Systems From Normal Operating Records," by T. P. Goodman, TRANS. ASME, vol. 79, 1957, pp. 567–575.

37   "Application of Correlation Analysis to the Detection of Periodic Signals in Noise," by Y. W. Lee, T. P. Cheatham, and J. B. Weisner, Proceedings of IRE, vol. 38, 1950.

38   "RC Memory Commutator for Signal-to-Noise Improvement," by C. I. Beard and E. N. Skomal, Review of Scientific Instruments, vol. 24, 1953.

APPENDIX   Brief Development of the Wiener-Hopf Equation

From Equation [95], the mean square error is defined as

$$(MSE) = \frac{1}{2T} \int_{-T}^{T} [q_{out} - q_{out(m)}]^2 dt \dots \dots [100]$$

where q_{out} is the calculated output and $q_{out(m)}$ is the measured output.

Since

$$q_{out} = \int_{0}^{\infty} [WF](t_c) q_{in}(t - t_c) dt_c \dots \dots \dots [101]$$

the mean square error is

$$(MSE) = \frac{1}{2T} \int_{-T}^{T} \left[\left(\int_{0}^{\infty} [WF](t_c) q_{in}(t - t_c) dt_c \right) - q_{out(m)} \right]^2 dt$$

$$\dots \dots [102]$$

It is required to choose the weighting function that makes the mean square error a minimum.   Let

$$\delta[WF] = \text{variation of the weighting function} \dots [103]$$

$$\delta(MSE) = \text{resultant variation of the mean square error} \dots [104]$$

Then

$$(MSE) + \delta(MSE) = \frac{1}{2T} \int_{-T}^{T} \left[\left(\int_{0}^{\infty} \{ [WF](t_c) \right. \right.$$

$$\left. \left. + \delta[WF](t_c) \} q_{in}(t - t_c) dt_c \right) - q_{out}(t) \right]^2 dt \dots \dots [105]$$

The variation of the mean square error is

$$\delta(MSE) = \frac{1}{2T} \int_{-T}^{T} \left[2 \int_{0}^{\infty} \delta[WF](t_{c1}) q_{in}(t - t_{c1}) dt_{c1} \right.$$

$$\int_{0}^{\infty} [WF](t_{c2}) q_{in}(t - t_{c2}) dt_{c2}$$

$$\left. - 2 q_{out(m)} \int_{0}^{\infty} \delta[WF](t_c) q_{in}(t - t_c) dt_c \right] dt \dots \dots [106]$$

Interchanging the order of the integration gives

$$\delta(MSE) = 2 \int_{0}^{\infty} \delta[WF](t_{c1}) dt_{c1} \int_{0}^{\infty} [WF](t_{c2}) dt_{c2}$$

$$\left[\frac{1}{2T} \int_{-T}^{T} q_{in}(t - t_{c1}) q_{in}(t - t_{c2}) dt \right]$$

$$- 2 \int_{0}^{\infty} \delta[WF](t_c) dt_c \left[\frac{1}{2T} \int_{-T}^{T} q_{out}(t) q_{in}(t - t_c) dt \right] \dots \dots [107]$$

Define

$$[ACF]_{q_{in}}(t_{c1} - t_{c2}) = \frac{1}{2T} \int_{-T}^{T} q_{in}(t - t_{c1}) q_{in}(t - t_{c2}) dt$$

$$= \text{input autocorrelation function} \dots \dots [108]$$

$$[CCF]_{[q_{in}; q_{out}]}(t_c) = \frac{1}{2T} \int_{-T}^{T} q_{out}(t) q_{in}(t - t_c) dt$$

$$= \text{input} - \text{output cross-correlation function} \dots \dots [109]$$

Introducing the correlation functions, the variation of the mean square error due to a variation of the weighting function is

$$\delta(MSE) = 2 \int_{0}^{\infty} \delta[WF](t_{c1}) dt_{c1}$$

$$\int_{0}^{\infty} [WF](t_{c2}) [ACF]_{q_{in}}(t_{c1} - t_{c2}) dt_{c2}$$

$$- 2 \int_{0}^{\infty} \delta[WF](t_c) [CCF]_{[q_{in}; q_{out}]}(t_c) dt_c \dots \dots [110]$$

The variation of the weighting function is zero for negative values of the argument t_c because the weighting function is so defined.   The variation of the mean square error is zero when

$$[CCF]_{[q_{in}; q_{out}]}(t_{c1}) = \int_{0}^{\infty} [WF](t_{c1}) [ACF]_{q_{in}}(t_{c1} - t_{c2}) dt_{c2} \dots [111]$$

The condition is valid only for $t_{c1} > 0$.

Discussion

Thomas P. Goodman.[8] The author has performed a valuable service in bringing together and critically comparing the various techniques which have been proposed for interpreting dynamic measurements. For the multicoupled operating system with random inputs, however, the solution has been carried farther than the author indicates. In reference (36) of the paper, a least-squares method is used to derive a multivariable Wiener-Hopf equation which gives the linear representation of a multicoupled

system which is optimum in a mean-square sense for the available data. In this optimum linear representation, as in the single-input case, the effects of uncertainties and signal corruptions, as well as the effects of nonlinearities, are optimally averaged out.[9]

The writer would be the first to agree that this linearized representation of a system is still inadequate for many purposes. Much more needs to be done, and the combination methods mentioned at the end of the paper appear to offer a promising approach.

[8] Assistant Professor of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Mass. Assoc. Mem. ASME.

[9] See also "Experimental Determination of System Characteristics From Correlation Measurements," by T. P. Goodman, ScD thesis, Dept. of Mechanical Engineering, Massachusetts Institute of Technology, 1955.

Solution of Reynolds' Equation for Finite Journal Bearings[1]

By O. PINKUS,[2] WEST LYNN, MASS.

Reynolds' differential equation for finite bearings was solved for journal bearings having 100 and 75-deg arcs, using a digital computer. These results were applied to partial bearings as well as to 3-groove and 4-groove full journal bearings. Results are given for L/D ratios ranging from $1^1/_2$ to $^1/_4$ and eccentricities up to 0.95. These results were correlated with previous results for different bearing arcs obtained by others and the author, to provide a comprehensive set of solutions of Reynolds' equation for finite bearings from a full 360-deg arc down to an arc of 75 deg. Expressions for calculating minimum film thickness, total oil flow, power loss, and temperature rise also are given.

NOMENCLATURE

The following nomenclature is used in the paper:

e = eccentricity, in.
h = film thickness, in.
j = power-loss coefficient, dimensionless
n = $2 e/C$, eccentricity ratio, dimensionless
p = pressure, psi
q = hydrodynamic flow coefficient, dimensionless
x, z = rectangular co-ordinates
C = diametral clearance, in.
D = diameter, in.
H = power loss, in-lb/sec
J = mechanical equivalent of heat, 778 ft-lb/Btu
L = length, in.
N = speed, rps
P = unit load, psi
S = Sommerfeld number, $(D/C)^2(ZN/P)$, dimensionless
T = temperature, deg F
U = linear speed, ips
W = bearing loading, lb
V = flux plot coefficient, dimensionless
Z = viscosity, lb-sec/in.^{2}
α = attitude angle, deg
φ = bearing arc, deg
ρ = specific weight, pci
c_p = specific heat, Btu/lb deg F
h_{min} = minimum film thickness, in.
p_l = inlet oil pressure, psi
q_0 = end flow coefficient, dimensionless
F_V = vertical force, dimensionless
F_H = horizontal force, dimensionless
Q_h = hydrodynamic oil flow, in.^{3}/sec
Q_0 = zero speed oil flow, in.^{3}/sec
θ_1 = leading edge of bearing, deg
θ_2 = trailing edge of bearing, deg

INTRODUCTION

Ever since Reynolds (1)[3] postulated the basic differential equation of hydrodynamic lubrication, a search went on for means and methods of solving this partial, nonhomogeneous equation for journal bearings. Most analytical attempts resorted to the radical approximation of an infinitely long bearing, and of these, Sommerfeld's (2) solution is the most prominent. This and similar solutions, however, were in striking disagreement with known facts and observation. It soon became clear that the behavior of an actual bearing cannot be represented, or even approximated, by a one-dimensional analysis. The prospects of obtaining analytical solutions of the intact Reynolds' equation were and remain dim, and investigators set out to obtain solutions by numerical means.

Following are the major milestones in the laborious field of obtaining solutions of the finite Reynolds' equation for journal bearings. Cameron and Wood (3) solved the equation for a full 360-deg bearing using Southwell's relaxation method. Needs (4), using the electrolytic tank analogy, solved Reynolds' equation for a 120-deg bearing arc. Wilcock and Rosenblatt (5), using a d-c network obtained solutions for a 150-deg arc. More recently, Walter and Sassenfeld (6) obtained solutions for both the 360 and 180-deg arcs by means of a modernized Gauss alogarithm. The author (7), using a dense grid of over 200 points in the bearing surface obtained solutions for the 150-deg arc on a digital computer which he also employed for the solution of elliptical bearings.

Thus in the course of years, more or less complete solutions were obtained for bearing arcs down to 120 deg. This paper extends the field of available solutions to the 100 and 75-deg arcs. Using these results, performance charts also were obtained for full 3-groove and 4-groove journal bearings which consist of three 100-deg arcs and four 75-deg arcs, respectively. The paper, by adding the previously obtained solutions to the present results, offers a comprehensive set of solutions of Reynolds' equation from 360 down to 75-deg arcs. These include eccentricity ratio, attitude angle, flow coefficients, and power-loss factor as functions of the Sommerfeld number. The present solutions, as well as all previous ones, are based on a load vector acting symmetrically with respect to the bearing boundaries.

SOLUTION OF REYNOLDS' EQUATION

The details of setting up Reynolds' equation in finite form with the inverse of the Sommerfeld number as the dependent variable, are given in reference (7). The major steps involve first the transformation of Reynolds' equation

$$\frac{\partial}{\partial x}\left(\frac{h^3}{Z}\frac{\partial p}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{h^3}{Z}\frac{\partial p}{\partial z}\right) = 6U\frac{\partial h}{\partial x} \quad \dots \dots [1]$$

[3] Numbers in parentheses refer to the Bibliography at the end of the paper.

into its dimensionless counterpart

$$\frac{\partial}{\partial x'}\left(h'^3\frac{\partial p'}{\partial x'}\right) + \left(\frac{D}{L}\right)^2 \frac{\partial}{\partial z'}\left(h'^3\frac{\partial p'}{\partial z'}\right) = 6\pi\frac{\partial h'}{\partial x'} \quad ..[2]$$

When Equation [2] is written in finite increments for each point of the grid in the bearing surface, the following expression is obtained

$$p_i = \frac{6\pi\frac{h_L - h_R}{\Delta x} + \left(\frac{D}{L}\right)^2\frac{h_T^3}{\Delta z^2}\ p_T + \frac{h_R^3}{\Delta x^2}\ p_R + \left(\frac{D}{L}\right)^2\frac{h_B^3}{\Delta z^2}\ p_B + \frac{h_L^3}{\Delta x^2}\ p_L}{\left(\frac{D}{L}\right)^2\frac{h_T^3}{\Delta z^2} + \frac{h_R^3}{\Delta x^2} + \left(\frac{D}{L}\right)^2\frac{h_B^3}{\Delta z^2} + \frac{h_L^3}{\Delta x^2}} \quad ...[3]$$

where subscripts L, T, R, B refer, respectively, to the left, top, right, and bottom points surrounding point p_i. For a grid of n points there will result a set of n algebraic equations in n unknowns.

The total number of discrete points used in obtaining solutions was approximately 200. The boundary points were located at a distance of $\Delta z/2$ and $\Delta z/2$ from the edges. Image points were assumed to insure zero boundary conditions. Otherwise, the ratio of $\Delta x/\Delta z$ was kept at 1. No negative pressure was allowed and, where zero pressure developed ahead of the trailing edge of the bearing arc, the remaining portion was assumed to be at zero pressure, which is the prevailing condition in actual bearings.

The n equations were solved simultaneously on a card program calculator following an iteration process with a built-in acceleration factor. A percentage error defined by $\Sigma(\Delta p/p)$ over the entire bearing surface was kept at a value of 0.004 insuring almost a perfect final answer.

With the pressure distribution solved, the vertical and horizontal load components were obtained from

$$\left.\begin{array}{l} F_V = \sum_1^{\dot{n}} p_i \cos\theta_i\ \Delta x\Delta z \\[1em] F_H = \sum_1^{\dot{n}} p_i \sin\theta_i\ \Delta x\Delta z \end{array}\right\} \quad ....[4]$$

The flow due to the hydrodynamic pressures at the edges of the bearing is given by

$$Q = \int \frac{h^3}{12\mu}\ \frac{\partial p}{\partial z}\bigg|_e\ dx = q\ \frac{\pi}{4}\ NDLC \quad ....[5]$$

from which

$$q = \frac{1}{3\pi}\left(\frac{D}{L}\right)^2 \sum_1^{\bar{}} h'^3 \frac{\Delta p'}{\Delta z'}\bigg|_e\ \Delta x' \quad ....[6]$$

the letter c denoting the appropriate boundary.

In the subsequent plots and equations, the value of unit loading P is defined as

$$P = \frac{W}{LD}$$

and not by the actual projected area of the partial bearing. This is merely a convention. The definition of unit loading could have been based on the actual projected area, and the numerical values of the Sommerfeld number would have been different. There is, however, some advantage in using LD as the reference area, because in making a comparison of various bearings from 360 down to 75 deg, it is desirable to keep the total load and reference area

the same. In that case the change in the value of the Sommerfeld number S in going from one value of φ to another represents the actual change in load capacity.

THE PARTIAL BEARING—100 AND 75-DEG ARCS

Fig. 1 shows the geometry of a partial bearing with the load vector acting through the center of the arc. The angle φ which is the angular span of the partial bearing is here called the bearing arc. As mentioned in the introduction, solutions were obtained for $\varphi = 100$ deg and $\varphi = 75$ deg for L/D ratios of $1\frac{1}{2}$, 1, $\frac{1}{2}$, and $\frac{1}{4}$, and eccentricity ratios up to 0.95.

Fig. 2 shows the locus of the shaft center for both the 100 and 75-deg arcs. These equilibrium positions were obtained after a series of trials in which an attitude angle α was sought, such that the resultant horizontal force would be zero. As in the case of a full bearing, the values of α for the different L/D ratios are not very far apart. (Table 3 in the next section gives the individual values of α for each condition.) In the polar plot of Fig. 2, only the case of $L/D = \frac{1}{4}$ shows a substantial deviation, the other cases blending into a single curve.

Figs. 3 and 4 give the relation between the eccentricity ratio and the Sommerfeld number for both bearing arcs. The case of $L/D = \infty$ is taken from Lee (8). These are the basic curves that determine load capacity and minimum film thickness. It should be pointed out again here that the unit loading is given by $P = W/LD$.
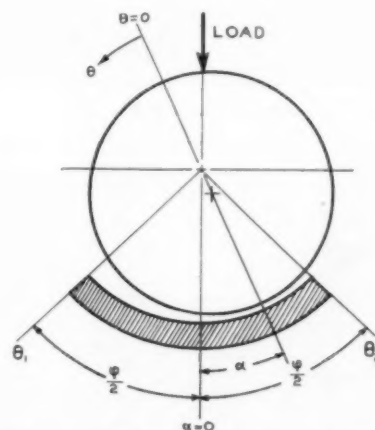
FIG. 1   GEOMETRY OF PARTIAL JOURNAL BEARING
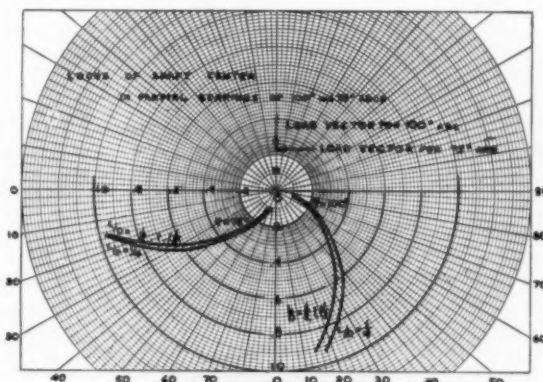
FIG. 2   LOCUS OF SHAFT CENTER

The flow of lubricant through a bearing, assuming zero inlet pressure, consists of two components, the shear and the hydrodynamic flows. The individual flow components are sketched in Fig. 5. The hydrodynamic flow which is due to the pressure gradients, exists at all four bearing boundaries when the pressure wave ends at θ_2; it exists only at three boundaries when the pressure wave ends ahead of θ_2. At the inlet edge some of the oil is immediately pumped back into the supply. This flow component is of no interest since it does not pass through the bearing and thus is not effective in cooling the bearing, and it also does not appear in the amount of lubricant flow as it is rejected back into the supply. Thus the total flow is made up of

$$\frac{\text{Side flow}}{\frac{\pi}{4} NDLC} = q = q_1 + q_2$$

and

$$\frac{\text{End flow}}{\frac{\pi}{4} NDLC} = q_0 = q_3 + q_4 = q_3 + (1 + n \cos \theta_2)$$

where q_1, q_2, and q_3 are the hydrodynamic factors as defined in Equation [5] and $q_4 = 1 + n \cos \theta_2$ is the shear flow. The total oil flow is thus given by

$$Q_A = (q + q_0) \frac{\pi}{4} NDLC \dots \dots \dots [7]$$

Figs. 6, 7, and 8 give the values of q and q_0 for $\varphi = 100$ deg and φ

FIG. 3 ECCENTRICITY RATIO FOR 100-DEG ARC
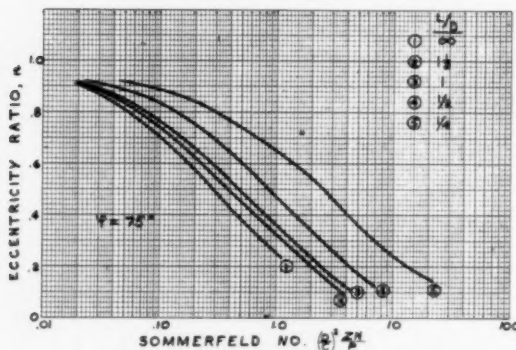
FIG. 4 ECCENTRICITY RATIO FOR 75-DEG ARC
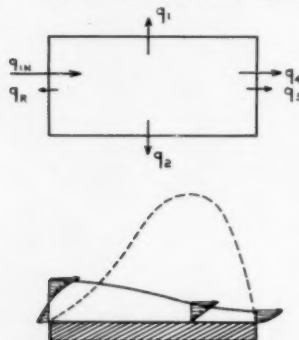
FIG. 5 LUBRICANT FLOW COMPONENTS
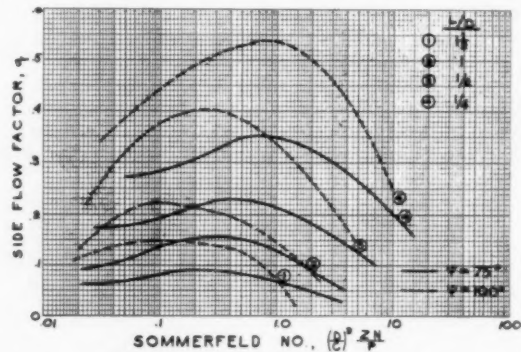
FIG. 6 SIDE FLOW IN PARTIAL BEARINGS
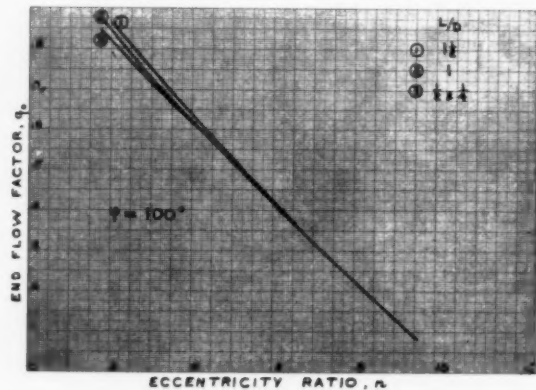
FIG. 7 END FLOW IN 100-DEG ARC
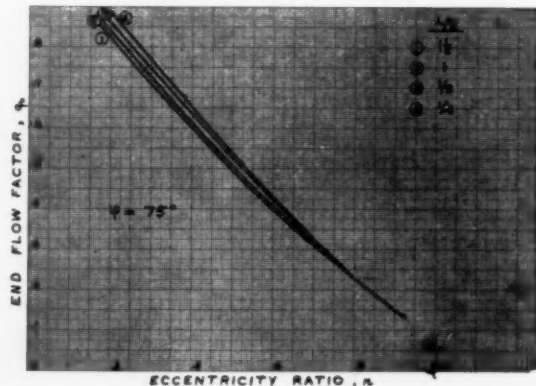
FIG. 8 END FLOW IN 75-DEG ARC

= 75 deg. Where the oil is delivered under pressure, an additional "zero speed" flow component exists. This, as developed in detail in references (5) and (7), is given by

$$Q_0 = \frac{p_1 C^3}{24Z} V \dots\dots\dots\dots\dots [8]$$

where $V = 31$ for $\varphi = 75$ deg and $V = 21$ for $\varphi = 100$ deg. The total oil flow through the bearing would then of course be

$$Q = (q + q_0) \frac{\pi}{4} NDLC + \frac{p_1 C^3 V}{24Z} \dots\dots\dots [9]$$

The power loss in a bearing can be written in differential form as

$$dH = \frac{ZU^2 RLd\theta}{h} = \frac{2ZU^2 RLd\theta}{C(1 + n \cos \theta)}$$

or

$$H = \frac{2ZU^2 RL}{C} \int_{\theta_1}^{\theta_2} \frac{d\theta}{1 + n \cos \theta}$$

This equation integrates into

$$H = \frac{2ZU^2 RL}{C} \frac{1}{(1 - n^2)^{1/2}} \tan^{-1} \frac{(1 - n^2)^{1/2} \sin \theta}{n + \cos \theta} \Big|_{\theta_1}^{\theta_2} \dots [10]$$

where

$$\theta_{2,1} = \pi \pm \frac{\varphi}{2} - \alpha$$

Equation [10] assumes a complete oil film in the bearing. This is nearly always the case for the two bearing arcs considered.

The Full Bearing—3 and 4 Groove

The previous solutions of Reynolds' equation can be utilized in obtaining performance data for full bearings which have three or four axial grooves. With the oil grooves subtending an angle of approximately 15 to 20 deg, a 3-groove bearing consists essentially of three 100-deg arcs and a 4-groove bearing of four 75-deg arcs. Previously (7), the author has applied the solution of the 150-deg arc to a full 2-groove bearing. In that case, the forces and flows

3-GROOVE BEARING

4-GROOVE BEARING

Fig. 9   Geometry of Axial-Groove Bearings

were identical for the partial and full bearing, but here, as shown in Fig. 9, at least two and oftentimes three lobes develop pressures affecting the load capacity and magnitude of flow. Thus an analysis has to include the forces present in all sectors. To get these values for the various attitudes involved, a generalized plot of F_v and F_A versus α for a given n (and a given L/D ratio) had first to be constructed. Then for a given attitude with respect to the lower lobe, the attitudes of the remaining lobes were established and all forces summed. This process of selecting various attitude angles had to be repeated until $\Sigma F_H = 0$. Then for this equilibrium α the values of F_v and q for all the lobes involved were summed algebraically.

Tables 1 and 2 give the basic results of the analysis. Fig. 10 gives a plot of the locus of the shaft center for a 2, 3, and 4-groove bearing. While in the case of the partial bearing the $\varphi = 75$-deg locus was closer to the vertical line than the $\varphi = 100$-deg case (Fig. 2), the locus of the full 4-groove bearing is considerably to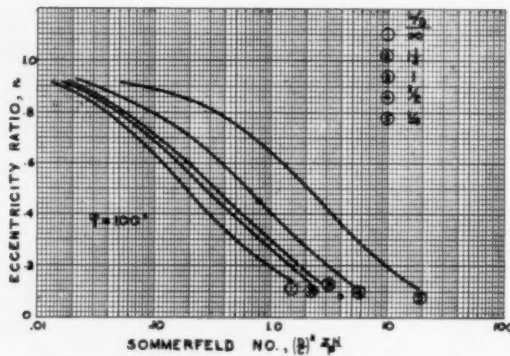 the right of the 3-groove bearing. This reflects the prominent effect that the bottom lobe in the 3-groove bearing, which has a wider arc than the 4-groove bearing, has on resisting the tendency of the

TABLE 1   THREE-GROOVE BEARING

L/D	n	α	S	q
$1/4$	0.2	74	9.61	0.3
	0.4	56	3.30	0.625
	0.6	41	1.17	0.90
	0.8	28	0.27	1.17
	0.9	20	0.0735	1.29
	0.95	14.5	0.024	1.35
$1/2$	0.2	72	3.0	0.245
	0.4	54	1.06	0.47
	0.6	39	0.385	0.68
	0.8	27	0.103	0.92
	0.9	20	0.033	1.04
	0.95	$15^1/_2$	0.012	1.08
1	0.2	69	1.33	0.16
	0.4	50	0.48	0.30
	0.6	37	0.187	0.425
	0.8	$26^1/_2$	0.059	0.51
	0.9	20	0.022	0.66
	0.95	15	0.009	0.715
$1^1/_2$	0.2	67	0.99	0.11
	0.4	49	0.37	0.205
	0.6	$36^1/_2$	0.15	0.29
	0.8	27	0.05	0.39
	0.9	$20^1/_2$	0.019	0.40
	0.95	15	0.0083	0.52

TABLE 2   FOUR-GROOVE BEARING

L/D	n	α	S	q
$1/4$	0.2	79	11.6	0.272
	0.4	72	4.5	0.56
	0.6	49	1.3	0.89
	0.8	25	0.23	1.15
	0.9	19	0.073	1.22
$1/2$	0.2	80	4.23	0.20
	0.4	72	1.7	0.42
	0.6	49	0.55	0.67
	0.8	25	0.11	0.84
	0.9	19	0.035	0.90
1	0.2	79	2.04	0.12
	0.4	72	0.87	—
	0.6	49	0.30	0.39
	0.8	25	0.07	0.51
	0.9	19	0.025	0.53
$1^1/_2$	0.2	80	1.9	0.081
	0.4	72	0.71	0.16
	0.6	49	0.245	0.29
	0.8	25	0.259	0.38
	0.9	$18^1/_2$	0.0222	0.37
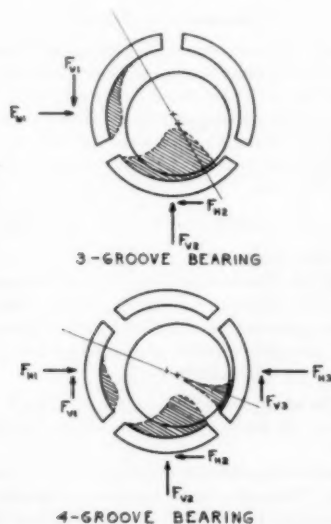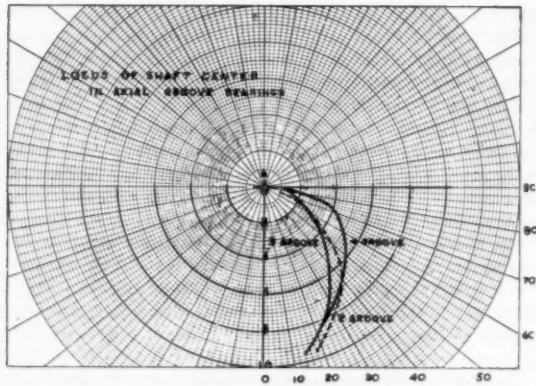
Fig. 10   Locus of Shaft Center in Axial-Groove Bearings

left lobe to force the shaft to the right. At small attitude angles this effect disappears.

Figs. 11 and 12 show the eccentricity ratio and flow coefficients. The hydrodynamic flow in full bearings consists obviously only of the side leakage as the lubricant discharged at the trailing edge is being recirculated and eventually discharged as side flow. Thus

$$Q_h = q \frac{\pi}{4} NDLC \dots \dots \dots [11]$$

The zero speed flow in axial-groove bearings is somewhat complicated. The pressure holes which find themselves in the con-
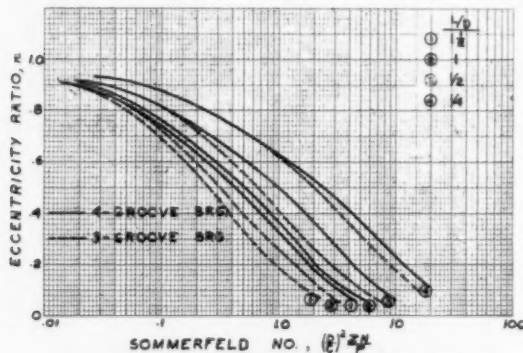
Fig. 11   Eccentricity Ratio in Axial-Groove Bearings
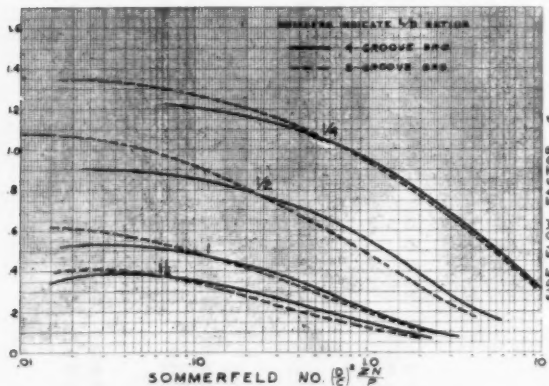
Fig. 12   Side Flow in Axial-Groove Bearings

verging film region are not able to admit much oil since the pressures there are at least equal to the inlet pressure. Thus in computing the zero speed flow for the 3-groove bearing, only one or two, depending on the attitude, of the oil grooves is admitting oil while in the 4-groove bearing only two grooves are operative. These conditions are reflected in the value of V in Equation (8)

3-groove bearing $\begin{cases} V = 21 & \text{for } \alpha > 50 \text{ deg} \\ V = 42 & \text{for } \alpha < 50 \text{ deg} \end{cases}$

4-groove bearing $V = 62$

The method of including the effect of the incomplete oil film on the power loss is described in references (5) and (7). A power-loss factor j is defined which includes the effect of both the attitude and incompleteness of the oil film, and the expression for power loss is given by

$$H = j \frac{2\pi^3 ZN^2 D^3 L}{C} \dots \dots \dots \dots [12]$$

where j is given in Fig. 13.
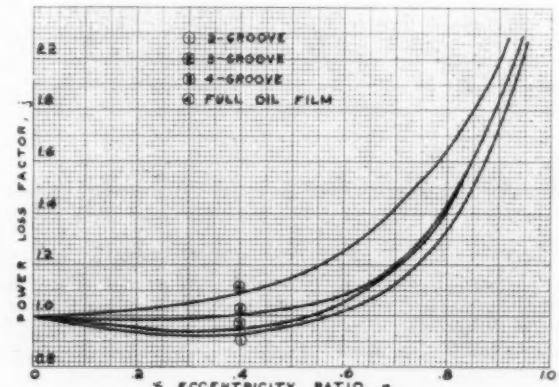
Fig. 13   Power Loss in Axial-Groove Bearings

Comprehensive Summary

As a comprehensive correlation, Table 3 gives a summary of the existing solutions for finite bearings. The results are taken from Cameron and Wood (3) for the 360-deg arc; from Walter and Sassenfeld (6) for the 180-deg arc; from Needs (4) for the 120-deg arc; and from the author's work for the 150, 100, and 75-deg arcs. No values for the end leakage are given for $\varphi > 120$ deg. This value is very close to

$$q_0 = 1 + n \cos (\pi + \varphi/2 - \alpha), \quad \varphi > 120 \text{ deg} \dots [13]$$

which represents just the shear-flow factor.

Figs. 14 and 15 show several plots of the Sommerfeld number as a continuous function of the bearing arc φ. Similar correlations and cross plots can be obtained for other cases.

Table 3, in grouping together a substantial body of data, suggests the following generalizations:

1   The value of α rises with a decrease in the L/D ratio and with a rise in φ. In the range of $1\frac{1}{2}$ to $\frac{1}{2}$ the variation of α with L/D is small.

2   The relative improvement in load capacity decreases with the L/D ratio. There is a larger increase in the relative load capacity in going from $L/D = \frac{1}{4}$ to $L/D = \frac{1}{2}$ than in going from an $L/D = 1$ to an $L/D = \infty$.

TABLE 3 PERFORMANCE DATA FOR FINITE JOURNAL BEARINGS

Sources

360 — Cameron & Woods (3)	180° — Walter & Sassenfeld (6)
150° — Pinkus (7)	120° — Needs (4)
100° — Pinkus	75° — Pinkus

n	α						S					
	360°	180°	150°	120°	100°	75°	360°	180°	150°	120°	100°	75°
L/D = ∞												
.2	68	61		53			.121	.181	.289	.432	.995	2.12
.4	62	50		39			.0628	.0884	.117	.181	.289	.397
.6	55	43		33			.0410	.0530	.0662	.0841	.117	.176
.8	43	33		30			.0224	.0261	.0265	.0340	.0398	.0530
.9	33	26					.0111	.0132	.0132		.0159	.0198
L/D = 1 1/2												
.2			62		58	57	.414	.418	.435		1.33	1.93
.4			51		40	40	.175	.182	.179		.454	.65
.6			42		32	31	.0875	.0935	.0943		.160	.213
.8			29		25	22	.0318	.0374	.0370		.0612	.064
.9			23		20	17	.0159	.0167	.0167		.0194	.023
L/D = 1												
.2	74	68	64	55	60	56	.627	.660	.714	1.00	1.33	2.33
.4	62	56	53	44	30	40	.255	.278	.275	.391	.202	.82
.6	50	45	45	36	31	31	.121	.129	.125	.164		.28
.8	36	32	28	31 1/2	26	22	.0446	.0460	.0410	.0550	.059	.077
.9	25	25	22		15	17	.0191	.0194	.0190		.023	.025
L/D = 1/2												
.2	75	71	64	61	61	56	2.03	2.05	2.10	2.42	3.12	4.27
.4	62	58	53	51	46	40	.788	.794	0.80	0.90	1.05	1.50
.6	46	45	42	37	32	31 1/2	.318	.320	0.31	.347	.395	.51
.8	33	31	24	32	26	22	.0786	.0927	.082	.097	.102	.135
.9	23	22	22		15	17	.0382	.0312	.031		.033	.036
L/D = 1/4												
.2	75	74	70	62	65	60	7.60	7.56	7.94	8.13	9.52	12.4
.4	61	59	57	53	49	42 1/2	2.64	2.64	2.86	2.86	3.33	4.0
.6	47	45	43	40	38	34	1.08	1.07	1.06	1.02	1.33	1.50
.8	31	31	29	36	26	23 1/2	.264	.265	.256	.232	.27	.31
.9	21	21	20		20	18	.0955	.0736	.074		.080	.077

n	q side					q * end		
	360°	150°	120°	100°	75°	120°	100°	75°
L/D = 1 1/2								
.2		.165		.01	.05		.87	.825
.4		.29		.135	.06		.65	
.6		.35		.145	.09		.43	.40
.8		.32		.14	.075		.23	.23
.9		.29		.115	.067		.13	.15
L/D = 1								
.2	.30	.27	.176	.135	.06	1.03	.65	.92
.4		.43	.224	.18	.14	.90		.61
.6		.56	.293	.21	.155	.75	.42	.41
.8	1.30	.46	.299	.215	.13	.54	.23	.23
.9		.43		.14	.093		.13	.15
L/D = 1/2								
.2	.37	.32	.238	.21	.135	1.1	.62	.86
.4	.45	.56	.360	.33	.20	1.0	.62	.64
.6	1.12	.715	.465	.34	.23	.91	.42	.42
.8	1.48	.725	.58	.38	.205	.84	.23	.23
.9		.695		.27	.18		.13	.15
L/D = 1/4								
.2	.40	.36	.265	.26	.18	1.1	.62	.88
.4	.79	.67	.406	.43	.28	1.1	.61	.675
.6	1.18	.86	.55	.53	.34	1.0	.41	.43
.8	1.57	.93	.78	.50	.33	1.0	.22	.23
.9		.81		.425	.28		.13	.15

* At $\varphi > 120$ the value of $q_0 \sim 1 + n\cos(\pi + \varphi/2 - \alpha)$.
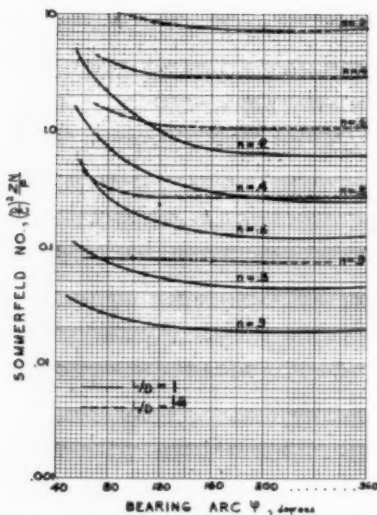
FIG. 14   LOAD CAPACITY AS A FUNCTION OF ARC AND ECCENTRICITY RATIO
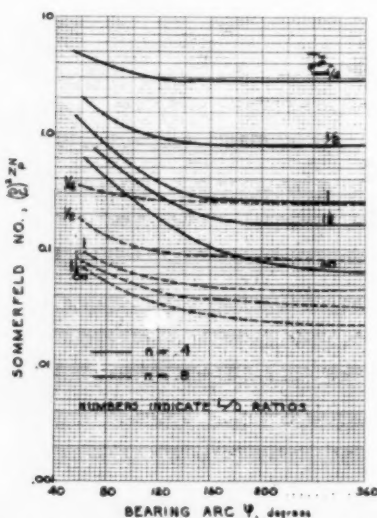
FIG. 15   LOAD CAPACITY AS A FUNCTION OF ARC AND L/D RATIO

3   There is a negligible change in load capacity with φ in the range of 360 to 150 deg.   This range extends with lower L/D ratios and higher n.   Thus at $n = 0.4$ and $L/D = 1\frac{1}{2}$ the curve of S versus φ begins to rise at $\varphi = 180$ deg; at $n = 0.8$, and $L/D = \frac{1}{4}$ the curve does not rise until a value of $\varphi = 120$ deg is reached.

4   There is an appreciable drop in the value of the side-leakage factor with a drop in φ.   As a function of n there is a maximum in the curve of q versus n.

PERFORMANCE CALCULATIONS

To summarize, following are the basic equations necessary to calculate the performance of partial or full bearings

$$n = f(\mathsf{S}) = f\left[\left(\frac{D}{C}\right)^2 \frac{Z_2 N}{P}\right]$$

$$e = \frac{Cn}{2}$$

$$h_{\min} = \frac{C}{2}(1 - n)$$

$$Q = (q + q_0)\frac{\pi}{4}NDLC + \frac{p_1 C^3 V}{24 Z_2}$$

$$H = \begin{cases} \dfrac{1}{2\pi(1-n^2)^{1/2}}\tan^{-1}\dfrac{(1-n^2)^{1/2}\sin\theta}{n+\cos\theta}\Big|_{\theta_1}^{\theta_2}\dfrac{2\pi^3 Z_2 N^2 D^3 L}{C} \\[2ex] j\,\dfrac{2\pi^3 Z_2 N^2 D^3 L}{C} \end{cases}$$

$$\Delta T = T_2 - T_1 = \frac{H}{12 J c_p \rho Q}$$

where $n = f(\mathsf{S})$, q, q_0, and j are to be taken from the appropriate charts and graphs. Z_2 is equal to the outlet oil viscosity.

In most expressions the viscosity Z appears as a parameter. Since the viscosity is a variable it has to be given some average, representative value.   References (5) and (7) deal with this problem to a greater extent and show that the outlet temperature supplies a reasonable approximation to such a representative viscosity.   The same references also outline a method of arriving at a correct answer without a prior knowledge of the value of Z_2.

ACKNOWLEDGMENT

BIBLIOGRAPHY

1   "On the Theory of Lubrication and Its Application to Mr. Beauchamp Tower's Experiments, Including Experimental Determination of the Viscosity of Olive Oil," by O. Reynolds, Philosophical Transactions of the Royal Society of London, England, vol. 177, 1887, pp. 157–234.

2   "Zur Hydrodynamische Theorie der Schmiermittel Reibung," by A. Sommerfeld, Zeitschrift für Mathematik und Physik, vol. 50, 1904, pp. 97–155.

3   "The Full Journal Bearing," by A. Cameron and Mrs. W. L. Wood, Proceedings of The Institution of Mechanical Engineers, vol. 161, 1949, pp. 59–64.

4   "Effects of Side Leakage in 120° Centrally Supported Journal Bearings," by S. J. Needs, Trans. ASME, vol. 56, 1934, p. 721.

5   "Oil Flow, Key Factor in Sleeve-Bearing Performance," by D. F. Wilcock and M. Rosenblatt, Trans. ASME, vol. 74, 1952, pp. 849–866.

6   "Gleitlagerberechnungen," by H. Sassenfeld and A. Walter, VDI Forschungsheft, vol. 20, no. 411, 1954.

7   "Analysis of Elliptical Bearings," by O. Pinkus, Trans. ASME, vol. 78, 1956, pp. 965–973.

8   "Analysis of Partial Journal Bearings Under Steady Load," by J. C. Lee, ASME Paper No. 55—LUB-1, unpublished.

Theoretical and Experimental Analysis of Hydrodynamic Gas-Lubricated Journal Bearings

By B. STERNLICHT[1] AND R. C. ELWELL,[2] SCHENECTADY, N. Y.

In recent years, gas-lubricated bearings have aroused a great amount of interest. The interest is centered on the advantages of gas bearings for four major applications: (a) High-temperature devices, for which there is yet no lubricant available. (b) Radioactive atmospheres, where conventional lubricants may break down. (c) Applications sensitive to contamination, such as jet engines, where fouling from lubricating oil becomes serious. (d) Low-friction devices—an especially important advantage because of the trend to high-speed machinery.

This paper presents a numerical solution for finite-width journal bearings and results of experiments conducted with air-lubricated hydrodynamic journal bearings. Comparison is made between theoretical and experimental results. Design formulas and recommendations for future studies also are included.

Nomenclature

THE following nomenclature is used in the paper:

A = coefficient (see Equation [12])
a = coefficient (see Equation [11])
C = diametral clearance, in.
D = journal-bearing diameter, in.
e = eccentricity, in.
F = dimensionless force
f = accelerating factor
h = film thickness, in.
L = axial length of bearing, in.
m = number of mesh subdivisions in θ or x-direction
n = number of mesh subdivisions in z-direction
N = rotational speed of journal, rps
P = pressure, psi
r = radius, in.
S = Sommerfeld number
U = relative velocity, ips
u, v, w = components of fluid velocity in x, y, z-direction, respectively
x, y, z = rectangular co-ordinates of a point
α = attitude angle, radians
γ = ratio of specific heat of gas lubricant
Δ = incremental distance
ϵ = eccentricity ratio
θ = angle, radians
μ = viscosity, lb-sec/in.^{2}

ρ = mass density, lb-sec^2/in.^{4}
ω = angular velocity, radians per sec

Subscripts

H = horizontal
i = index defining the value of z in the journal bearing running from 1 to n
j = index defining the value of x or θ in the journal bearing running from 1 to m
k = number of iterations
V = vertical
avg = average
0 = ambient

Superscript

$'$ = prime is dimensional

Introduction

Air-lubricated bearings have a long history. Journal, thrust, and spherical designs have been built. They have been lubricated both by normal hydrodynamic action and by the hydrostatic introduction of air at elevated pressures. A. Kingsbury (1)[3] constructed a hydrodynamic journal bearing in 1896 and conducted a series of tests to determine its characteristics. H. F. Bruback (2) and H. Drescher (3) have used hydrodynamic journal and thrust bearings for various applications. Fuller (4) has discussed the frictional properties of air bearings. The phenomenon of instability of hydrodynamic air-lubricated journal bearings has been discussed in references (3, 5, 6).

In 1913, W. J. Harrison (7) developed an air-bearing theory for infinite-width bearings. Using numerical methods, he obtained a solution of the nonlinear differential equation. The analysis assumes isothermal expansion and compression of the gas.

J. S. Ausman (8) relaxed Harrison's (7) isothermal restriction and, using perturbation methods, obtained an approximate solution for gas-lubricated bearings. Solutions were presented for infinite-width bearings at small eccentricity ratios.

A paper by M. Wildmann (6) presented quantitative data from experiments conducted on gas-lubricated hydrodynamic journal bearings. Test results for air, neon, and helium lubrication were presented.

Theoretical Analysis

The following approximations are made in the Navier-Stokes equation:

1 Inertia terms are neglected with respect to pressure-gradient terms.

2 Vertical fluid velocity v is neglected with respect to horizontal fluid velocities.

3 The predominant viscous shear stresses are $\partial^2 u/\partial y^2$ and $\partial^2 w/\partial y^2$. All other viscous shear stresses are negligible by comparison.

[1] Engineering Specialist, General Engineering Laboratory, General Electric Company. Assoc. Mem. ASME.
[2] Engineer, Fluid Film Bearings, General Engineering Laboratory, General Electric Company. Assoc. Mem. ASME.

[3] Numbers in parentheses refer to the Bibliography at the end of the paper.

4   Pressure is independent of the y-co-ordinate.
5   Density is independent of the y-co-ordinate.
6   The fluid is Newtonian.

Neglecting the effect of curvature, it is possible to analyze the circular bearing using the configuration shown in Fig. 1, where

h' = film thickness
P' = pressure at any point
x' = co-ordinate in direction of flow
z' = width co-ordinate
μ' = fluid viscosity
U' = relative velocity
ρ' = density at any point
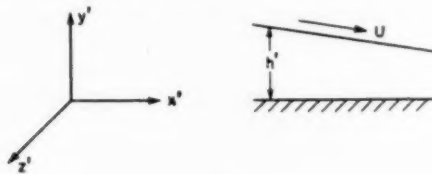
Fig. 1
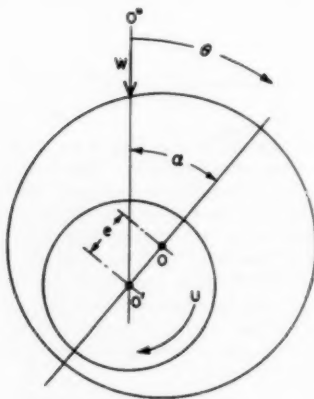
Fig. 2

The film thickness h' in the bearing with certain eccentricity e is calculated approximately from the following Equation (see Fig. 2)

$$h' = \frac{C'}{2} + e \cos(\theta - \alpha) \dots\dots\dots\dots [1]$$

where

O is the bearing center
O' is the shaft center
C' = diametral clearance
e' = eccentricity
α = attitude angle
θ = angle measured from vertical load on bearing to h'
ϵ = eccentricity ratio = $2e'/C'$

Substituting the parameter $h = h'/C$, Equation [1] is made dimensionless

$$\therefore h = \frac{1}{2} + \frac{1}{2}\,\epsilon \cos(\theta - \alpha) \dots\dots\dots\dots [2]$$

From the Navier-Stokes equation and the continuity equation for compressible fluid flow we get

$$\frac{\partial}{\partial x'}\left[\rho'\,\frac{h'^3}{\mu'}\,\frac{\partial P'}{\partial x'}\right] + \frac{\partial}{\partial z'}\left[\frac{\rho'h'^3}{\mu'}\,\frac{\partial P'}{\partial z'}\right] = 6U\,\frac{\partial}{\partial x'}\,[\rho'h'] \dots [3]$$

To write Equation [3] in dimensionless form, the following parameters are used

$$
\left.
\begin{aligned}
&x = \frac{x'}{D'} \qquad D'\,\partial x = \partial x' \\[4pt]
&z = \frac{z'}{L'} \qquad L'\,\partial z = \partial z' \\[4pt]
&h = \frac{h'}{C'} \\[4pt]
&\rho = \frac{\rho'}{\rho_0'} \\[4pt]
&\mu = \frac{\mu'}{\mu_{avg}'} \\[4pt]
&U' = \pi D'N' \\[4pt]
&P = \frac{P'}{\mu_{avg}'N'}\left(\frac{C'}{D'}\right)^2 \qquad \mu_{avg}'N'\left(\frac{D'}{C'}\right)^2 \partial P = \partial P'
\end{aligned}
\right\} \dots [4]
$$

Substituting [4] in Equation [3] we get

$$\frac{\partial}{D'\partial x}\left[\rho_0'\rho\,\frac{(C'h)^3}{\mu\mu_{avg}'}\,\mu_{avg}'N'\left(\frac{D'}{C'}\right)^2\frac{\partial P}{D'\partial x}\right]$$
$$+ \frac{\partial}{L'\partial z}\left[\rho\rho_0'\,\frac{C'h^3}{\mu\mu_{avg}'}\,\mu_{avg}'N'\left(\frac{D'}{C'}\right)^2\frac{\partial P}{L'\partial z}\right]$$
$$= 6\pi D'N'\frac{\partial}{D'\partial x}\,[\rho_0'\rho hC'] \dots\dots\dots [5]$$

which reduces to

$$\frac{\partial}{\partial x}\left[\frac{\rho h^3}{\mu}\,\frac{\partial P}{\partial x}\right] + \left(\frac{D}{L}\right)^2\frac{\partial}{\partial z}\left[\frac{\rho h^3}{\mu}\,\frac{\partial P}{\partial z}\right] = 6\pi\frac{\partial}{\partial x}\,(\rho h) \dots [6]$$

For the incompressible case, the density remains constant. Therefore, ρ is canceled in each term of Equation [6].

Referring to Fig. 3 and writing Equation [6] in difference form, we get

Fig. 3

$$\frac{\partial}{\partial x}\left[\frac{\rho h^3}{\mu}\frac{\partial P}{\partial x}\right] = \frac{\left(\frac{h^3\rho}{\mu}\right)_{i,j+\frac{1}{2}}\left(\frac{P_{i,j+1}-P_{i,j}}{\Delta x}\right) - \left(\frac{h^3\rho}{\mu}\right)_{i,j-\frac{1}{2}}\left(\frac{P_{i,j}-P_{i,j-1}}{\Delta x}\right)}{\Delta x}$$

$$\frac{\partial}{\partial z}\left[\frac{\rho h^3}{\mu}\frac{\partial P}{\partial z}\right] = \frac{\left(\frac{h^3\rho}{\mu}\right)_{i+\frac{1}{2},j}\left(\frac{P_{i+1,j}-P_{i,j}}{\Delta z}\right) - \left(\frac{h^3\rho}{\mu}\right)_{i-\frac{1}{2},j}\left(\frac{P_{i,j}-P_{i-1,j}}{\Delta z}\right)}{\Delta z}$$

$$\frac{\partial}{\partial x}(\rho h) = \frac{(\rho h)_{i,j+\frac{1}{2}} - (\rho h)_{i,j-\frac{1}{2}}}{\Delta x}$$

$\qquad\qquad\qquad\qquad$ .... [7]

Recombining the terms, Equation [6] as a difference equation is

$$\frac{\left(\frac{h^3\rho}{\mu}\right)_{i,j+\frac{1}{2}}\left(\frac{P_{i,j+1}-P_{i,j}}{\Delta z}\right) - \left(\frac{h^3\rho}{\mu}\right)_{i,j-\frac{1}{2}}\left(\frac{P_{i,j}-P_{i,j-1}}{\Delta x}\right)}{\Delta x}$$

$$+ \left(\frac{D'}{L'}\right)^2 \frac{\left[\left(\frac{h^3\rho}{\mu}\right)_{i+\frac{1}{2},j}\left(\frac{P_{i+1,j}-P_{i,j}}{\Delta z}\right) - \left(\frac{h^3\rho}{\mu}\right)_{i-\frac{1}{2},j}\left(\frac{P_{i,j}-P_{i-1,j}}{\Delta z}\right)\right]}{\Delta z} = 6\pi\frac{\left[(\rho h)_{i,j+\frac{1}{2}} - (\rho h)_{i,j-\frac{1}{2}}\right]}{\Delta x} \quad \ldots\ldots [8]$$

The pressure at the center of any grid is

$$P_{i,j} = \frac{6\pi\frac{\left[(\rho h)_{i,j-\frac{1}{2}} - (\rho h)_{i,j+\frac{1}{2}}\right]}{\Delta x} + \left(\fra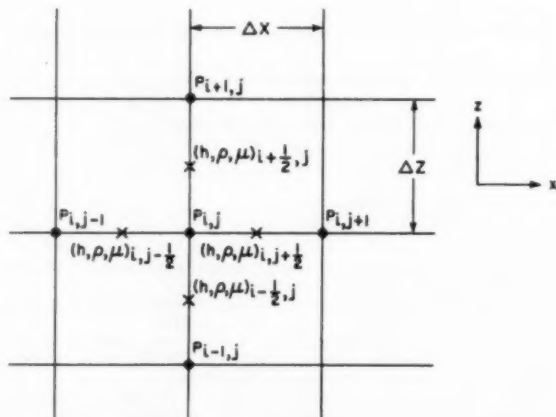c{h^3\rho}{\mu}\right)_{i,j+\frac{1}{2}}\frac{P_{i,j+1}}{\Delta x^2} + \left(\frac{h^3\rho}{\mu}\right)_{i,j-\frac{1}{2}}\frac{P_{i,j-1}}{\Delta x^2}}{\left(\frac{h^3\rho}{\mu}\right)_{i,j+\frac{1}{2}}\frac{1}{\Delta x^2} + \left(\frac{h^3\rho}{\mu}\right)_{i,j-\frac{1}{2}}\frac{1}{\Delta x^2} + \left(\frac{D'}{L'}\right)^2\left[\left(\frac{h^3\rho}{\mu}\right)_{i+\frac{1}{2},j}\frac{1}{\Delta z^2} + \left(\frac{h^3\rho}{\mu}\right)_{i-\frac{1}{2},j}\frac{1}{\Delta z^2}\right]}$$

$$+ \frac{\left(\frac{D'}{L'}\right)^2\left[\left(\frac{h^3\rho}{\mu}\right)_{i+\frac{1}{2},j}\frac{P_{i+1,j}}{\Delta z^2} + \left(\frac{h^3\rho}{\mu}\right)_{i-\frac{1}{2},j}\frac{P_{i-1,j}}{\Delta z^2}\right]}{\left(\frac{h^3\rho}{\mu}\right)_{i,j+\frac{1}{2}}\frac{1}{\Delta x^2} + \left(\frac{h^3\rho}{\mu}\right)_{i,j-\frac{1}{2}}\frac{1}{\Delta x^2} + \left(\frac{D'}{L'}\right)^2\left[\left(\frac{h^3\rho}{\mu}\right)_{i+\frac{1}{2},j}\frac{1}{\Delta z^2} + \left(\frac{h^3\rho}{\mu}\right)_{i-\frac{1}{2},j}\frac{1}{\Delta z^2}\right]}$$

$\qquad\qquad\qquad\qquad$ ... [9]

Both density and viscosity in Equation [9] are functions of temperature and pressure. Therefore, in order to obtain a rigorous solution for the gas lubricated bearing, the energy equation and the equation of state must be simultaneously solved with the Reynolds equation. This solution of the problem was programmed for the IBM 650 digital computer and will be reported on in a later paper. This method of solution, even with the aid of a high speed computer, is still very time consuming. A simpler method of solution was sought; it would not be quite as rigorous as the method just mentioned, but considerably less time consuming. One simplified method, assuming isothermal conditions in theoretical analysis, yielded an unsatisfactory correlation with experimental results. (This point was confirmed in reference 8.) When a more general pressure-density relationship (Equation [10]) was used and viscosity was considered to be constant, a closer correlation between theory and experiments was found. Physically this may be attributed to the cancellation of the conduction term and the viscous shear term. This semiempirical method was of great interest for it reduced the time of computation by approximately sixty per cent as compared with the more rigorous analysis. This paper, therefore, neglects viscosity changes and related pressure to density by using Equation [10].

$$\rho' = \rho_0'\left(\frac{P_0' + P'}{P_0'}\right)^{\frac{1}{\gamma}} \qquad \ldots\ldots\ldots\ldots [10]$$

where P' is pressure above or below ambient.

This written in dimensionless form becomes

$$\rho = \frac{\rho'}{\rho_0'} = \left[1 + \frac{P'}{P_0'}\mu_{\text{avg}}'N'\left(\frac{D'}{C'}\right)^2\right]^{\frac{1}{\gamma}}$$

For the isothermal case, the relation is simplified, for the pressure is proportional to the density ($P' \propto \rho'$).

For the journal bearing under investigation there is no film-thickness variation in the axial, or z-direction, hence

$$h_{i+\frac{1}{2},j} = h_{i-\frac{1}{2},j}$$

Also, it is assumed that the viscosity is constant throughout the field. Hence: $\mu_{i+\frac{1}{2},j} = \mu_{i-\frac{1}{2},j} = \mu_{i,j+\frac{1}{2}} = \mu_{i,j-\frac{1}{2}} = 1$

Thus Equation [9] can be written in the following form

$$P_{i,j} = a_0 + a_1 P_{i+1,j} + a_2 P_{i,j+1} + a_3 P_{i-1,j} + a_4 P_{i,j-1} \ldots [11]$$

where

$$a_0 = \frac{6\pi \left[(\rho h)_{i,j-\frac{1}{2}} - (\rho h)_{i,j+\frac{1}{2}} \right]}{A \Delta x}$$

$$a_1 = \left(\frac{D'}{L'} \right)^2 \frac{(h^3\rho)_{i+\frac{1}{2},j}}{A \Delta z^2}$$

$$a_2 = \left(\frac{D'}{L'} \right)^2 \frac{(h^3\rho)_{i-\frac{1}{2},j}}{A \Delta z^2}$$

$$a_3 = \frac{(h^3\rho)_{i,j+\frac{1}{2}}}{A \Delta x^2}$$

$$a_4 = \frac{(h^3\rho)_{i,j-\frac{1}{2}}}{A \Delta x^2}$$

$$A = \left(\frac{D'}{L'} \right)^2 \frac{(h^3\rho)_{i+\frac{1}{2},j} + (h^3\rho)_{i-\frac{1}{2},j}}{\Delta z^2}$$
$$+ \frac{1}{\Delta x^2} \left[(h^3\rho)_{i,j+\frac{1}{2}} + (h^3\rho)_{i,j-\frac{1}{2}} \right]$$

$$\quad\quad [12]$$

Solution of Equation [11] is performed in the following manner:

1 L/D, ϵ, m, and n are given, and α is assumed.

2 The value of the film thickness at every point is determined from Equation [2].

3 $P_{i,j}$ is assumed equal to zero throughout the field. This means that density throughout the field is unity from Equation [10].

4 Having h and $P_{i,j}$, the first approximation to the pressure field is determined from Equation [11].

5 This pressure field is improved several times by iteration.

6 The iterated pressure field is now used to calculate the density distribution from Equation [10].

7 Another pressure field can now be calculated from Equation [11] by using h as calculated in step 2, ρ calculated in step 6, and the last iterated pressure $P_{i,j}$ from step 5.

8 This cycle of pressure and density iteration is continued until $P_{i,j}$ reaches the tolerance factor specified for the run

$$\text{Tolerance factor} = \frac{\sum_{j=1}^{m} \sum_{i=1}^{n} \left| (P_{i,j})_k - (P_{i,j})_{k-1} \right|}{\sum_{j=1}^{m} \sum_{i=1}^{n} \left| (P_{i,j})_k \right|}$$
$$< 0.1 \text{ per cent} \dots \dots [13]$$

9 The final pressure field is then used to compute the vertical and horizontal components of force.

The force at a point is equal to the pressure at that point times the area. The total force on the bearing is the integral of pressure times area summed over the whole bearing

$$\left. \begin{array}{l} F_V = - \sum_{j=1}^{m} \sum_{i=1}^{n} (P_{i,j})_k \cos \theta \, \Delta x \Delta z \\[2mm] F_H = \sum_{j=1}^{m} \sum_{i=1}^{n} (P_{i,j})_k \sin \theta \, \Delta x \Delta z \simeq 0 \end{array} \right\} \dots [14]$$

If the horizontal component of force is not equal to zero, a new α is assumed and steps 1–8 are repeated until $F_H \simeq 0$. This satisfies the equilibrium of force.

In order to accelerate the process of iterations, an accelerating factor f is employed. This factor may vary from row to row or from iteration to iteration. The value chosen at any time is based on experience. The following equation is used for the adjustment of pressure between successive iterations

$$P_{i,j} = f[(P_{i,j})_k - (P_{i,j})_{k-1}] + (P_{i,j})_{k-1} \dots \dots [15]$$

To give some idea of the cost of using a computer such as the IBM 650 for this calculation, each case solved ($m \times n = 12 \times 12$) required 20 to 25 iterations and between 20 to 30 min. At an hourly rate of $70 to $80, this means a cost of $25 to $40 per case. It should be remembered, however, that this does not include the cost of developing the program, which may take a matter of one or two man-months.

Once the pressure profile is determined, it is possible to calculate the frictional force and flow.

Fig. 4 shows a typical pressure profile obtained analytically for $L/D = 1.5$ and $\epsilon = 0.4$. The pressures are expressed in dimensionless form and are calculated circumferentially at sections 15 per cent and 45 per cent of the bearing length. Note that pressures below atmospheric exist in the diverging region and that positive pressure extends over an arc greater than 180 deg.
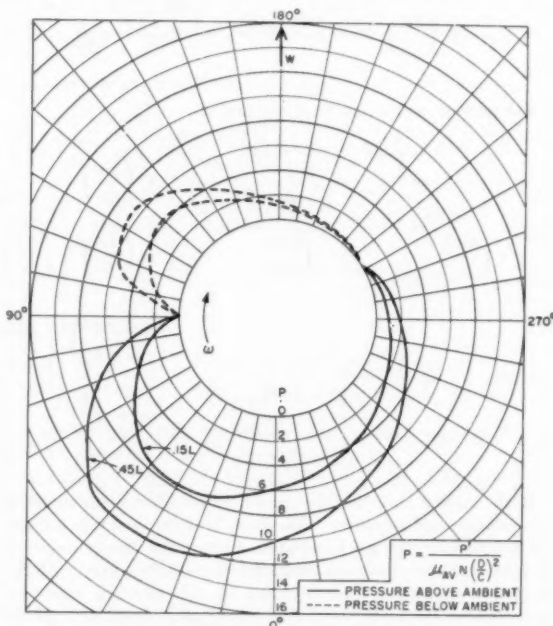
Fig. 4 Theoretical circumferential pressure distribution for $L/D = 1.5$ and $\epsilon = 0.4$

Fig. 5 shows Sommerfeld number versus eccentricity ratio for $L/D = 1$ and 1.5. Similar curves can be obtained for other L/D ratios. Comparison is also made with the incompressible case. Note that the Sommerfeld number for the compressible case $L/D = 1$ is comparable to the incompressible case $L/D = 1.5$. The compressible case yields approximately 30 per cent higher load than the incompressible case.

Fig. 6 shows the eccentricity loci of gas-lubricated journal bearings for $L/D = 1$ and 1.5. When these curves are compared to the incompressible case, it becomes apparent that, for the same eccentricity ratio, the attitude angle for the compressible case is larger than for the incompressible case. Further, when a comparison is made of the compressible solution with laminar and
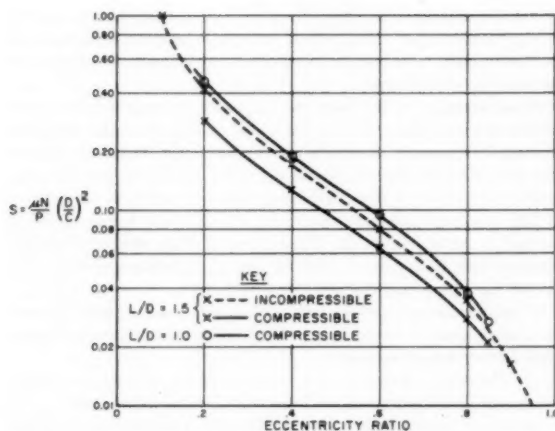
Fig. 5   Theoretical Sommerfeld number versus eccentricity ratio
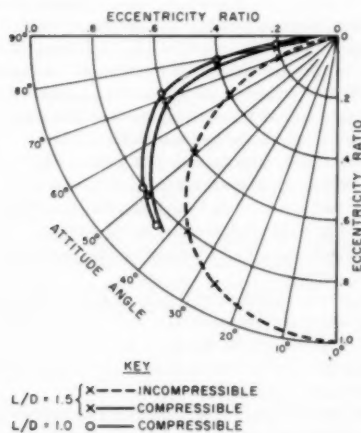
Fig. 6   Theoretical eccentricity ratio versus attitude angle

turbulent incompressible cases, agreement is better with the turbulent condition (9).

Equation [3] can be expressed in a different form if other parameters are used in Equation [4].   Using the following parameters

$$\left.\begin{array}{l} x' = r\theta \\ z' = rz \\ U^* = \pi\, D'N' \\ \rho' = \rho\rho_0' \\ h' = hC' \end{array}\right\} \quad \dots\dots\dots\dots [16]$$

and substituting them into Equation [3] we get

$$\left.\begin{array}{l} \dfrac{\partial}{\partial\theta}\left[\rho h^3\,\dfrac{\partial P'}{\partial\theta}\right] + \dfrac{\partial}{\partial z}\left[\rho h^3\,\dfrac{\partial P'}{\partial z}\right] \\[2mm] \qquad\qquad = 3\pi \dot{P_0}'\,\dfrac{\mu'N'}{P_0'}\left(\dfrac{D'}{C'}\right)^2 \dfrac{\partial}{\partial\theta}\,[\rho h] \end{array}\right\} \dots [17]$$

$$\dfrac{\partial}{\partial\theta}\left[\rho h^3\,\dfrac{\partial P'}{\partial\theta}\right] + \dfrac{\partial}{\partial z}\left[\rho h^3\,\dfrac{\partial P'}{\partial z}\right] = 3\pi P_0'\,\mathsf{S}\,\dfrac{\partial}{\partial\theta}\,[\rho h]$$

Note that the Sommerfeld number uses ambient pressure rather than average pressure over the projected area.   This equation was used only for comparison of our results with those reported in reference (8).

Description of test apparatus

In general, the equipment consists of five major items:

1   Precision spindle with one journal at each end.
2   Variable-speed drive motor.
3   Test bearings.
4   Loading equipment.
5   Instrumentation for measuring the following quantities: (a) Speed; (b) pressure; (c) fluid-film thickness; (d) angle; and (e) contact between journal and bearing.   Figs. 7 and 8 show the test setup with the latter giving a clearer view of the instrumentation and loading arrangement.

A detailed description of each piece of equipment follows:

Spindle.   A high-precision spindle with special shaft extensions was employed.   The extension in one end constituted the $2^1/_2$-in. journal and that at the other end was the 2-in. journal. Both journal diameters were ground to 5-microin-rms surface finishes.   Diametral tolerances were $+0.0000$ to -0.0001 in. In addition, total maximum runout on either journal was held to less than 0.00004 in.   Journal material was AISI 416 stainless steel of 25 Rockwell C hardness.   The spindle was belt driven by a variable-speed motor.

Drive Motor.   The drive motor was a $1/_2$-hp General Electric Thy-mo-trol unit with speed adjustment from 86 to 1725 rpm. Higher speed on the test journal was obtained through a pulley ratio between motor and spindle.

Test Bearings.   The test bearings were of 2 and $2^1/_2$-in. diam. L/D ratio for both was 1.5.   The bores were ground to better than 5-microin-rms surface finish.   After completion of the first series of tests, the bearing bores were ground to give greater clearance.

Bore diameters were measured in a temperature-controlled
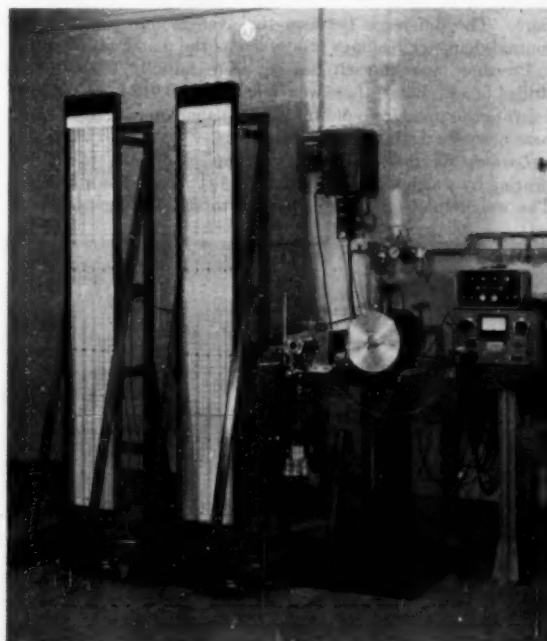
Fig. 7   Setup of test apparatus
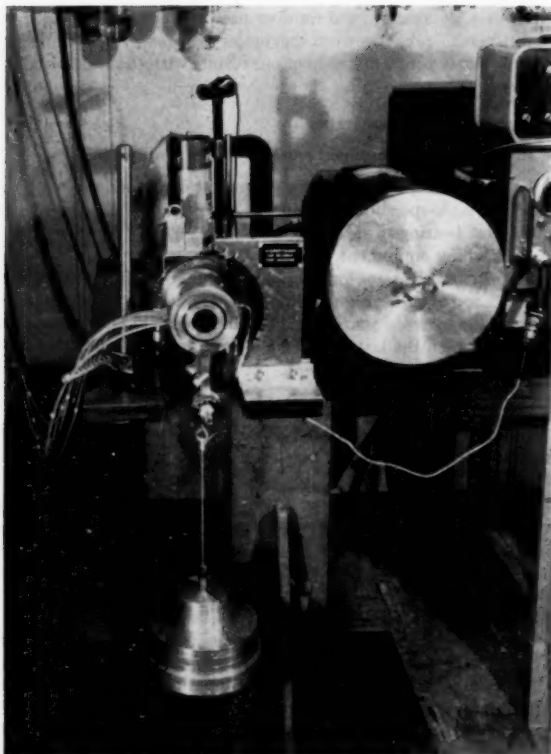
Fig. 8 Test setup showing instrumentation and loading arrangement

room with a Pratt and Whitney bore checker, having a smallest reading of 10 microin. Ninety-six readings were taken on each bore. The difference between the average bore readings and journal-diameter readings was taken as the diametral clearance.

Pressure taps in each bearing were initially 0.005-in-diam drilled holes. Later, these were enlarged to 0.012-in. diam to accelerate the pressure build-up in the manometers, Fig. 9. Both bearings were made of SAE 660 bronze.

Loading Equipment. Load was applied directly to the test bearing by weights on the special weight hanger shown in Fig. 8. The weight platform was connected to the bearing by a yoke.

A ball bearing on each arm of the yoke ran in a track on each end of the test bearing. This allowed the test bearing to be rotated without moving the load and insured that the load would always be directly downward.

Axial position of the load was adjusted by means of a collar that slid along the yoke. Axial alignment of the load could be obtained extremely accurately from observation of the axial-pressure distribution. For true axial alignment, the pressure distribution must be symmetrical about the central axis.

The bearing was restrained from rotation by a torque arm attached to the bearing wall. The other end of the arm was held by a stand with slots spaced every 10 deg. The bearing could be indexed through a full 360 deg.

Instrumentation. (a) *Speed.* A frequency meter and counter indicated speed to an accuracy of ±0.2 per cent. The meter was excited by a magnetic pickup.

(b) *Pressure.* Pressure was read on 72-in. manometers using water for lower pressures, and mercury for the higher ones. The accuracy of measurement was ±0.02 psi. The 2-in. bearing had seven holes equally spaced along its length, and the $2^{1}/_{2}$-in. bearing had nine holes. These holes were drilled for the purpose of measuring pressure distribution. By rotating the bearing through 360 deg it was possible to obtain the circumferential pressure distribution also. Thin-walled plastic tubing connected the manometers to pressure fittings in the wall of the bearing.

(c) *Fluid-film thickness.* A special capacitance pickup was mounted in the wall of the bearing so that the pickup face was about 0.001 in. beneath the bearing surface, Fig. 9.

The output of this pickup was put into a Fielden proximity meter where it was amplified and indicated by a meter.

This meter had a 4-in. scale length and 50 divisions on the scale. It was normally used with sensitivity of 0.001 in. for full-scale deflection. The accuracy of reading the film thickness was therefore within 5 microin.

The pickup was calibrated by jacking the bearing on the shaft and reading bearing movement on a Pratt and Whitney Electrolimit Gage, Fig. 10. This instrument had a full-scale range of 0.001 in. with 20-microin. subdivisions. Accuracy of reading this gage is also within 5 microin.

The Pratt and Whitney gage was checked against gage blocks accurate to 4 microin. and the necessary correction applied to the gage readings during calibration of the pickup. In operation, the bearing was indexed to locate the pickup in a desired position.

(d) *Angle.* A plastic disk with angles marked every 5 deg was pressed on the end of the test bearing. This was referred to index

Fig. 9 Special capacity pickup mounted in wall of bearing
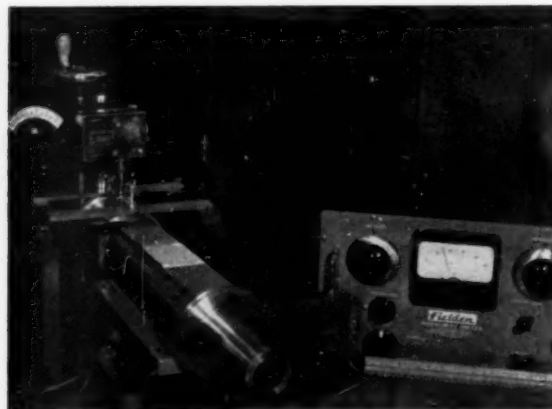
Fig. 10 Pratt and Whitney Electrolimit Gage used for calibrating bearing movement

slots cut in the test stand, Fig. 8. Thus the circumferential pressure distribution, film thickness, and attitude angle could be obtained by indexing the bearing through 360 deg in intervals of 10 deg. The accuracy of this measurement is estimated to be ±3 deg.

(e) *Contact between shaft and bearings.* This was determined by a simple electrical continuity circuit. If the shaft touched the bearing, a light would come on. When only asperities were touching, brightness of the light would give an indication of the magnitude of contact area, Fig. 8.

Test Procedure

1 The bearing was loaded to about 2 psi at standstill. This was done to avoid instability under operation.

2 The journal and bearing were wet completely with a fluid which vaporized and did not leave a residue.

3 The motor was started and accelerated to the desired test speed.

4 It was then run for several minutes until the fluid vaporized and the torque reduced to a steady-state value.

5 The desired load was then applied.

6 The bearing was indexed to locate the pressure taps or film-thickness gage at the desired angle.

7 Before pressure was read, the manometers were allowed to reach a steady reading. This was done by keeping a pressure-time graph to see when the pressure leveled off.

8 When film thickness was to be measured, the film-thickness gage was read and the bearing was indexed to the next angle.

9 Steps 7 and 8 were repeated at the various angles. These data gave axial and circumferential pressure distribution, film thickness, and attitude angle for a specific speed and load.

10 Shutdown procedure:

(a) The load was removed except for about 2 psi.
(b) The speed was reduced to approximately 2000 rpm.
(c) Liquid was injected into the bearing in the vacuum region.
(d) The motor was then shut off.
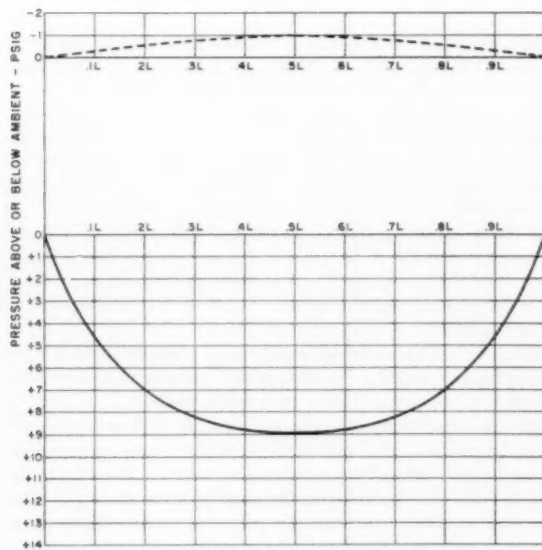
Experimental Results

The experimental data, obtained as described previously, are presented in graphical form in Figs. 11 through 16. Fig. 11 shows a typical pressure distribution for a specific load, speed, and bearing geometry. Both here and in the theoretical results, Fig. 4, the positive pressure distribution extends over more than 180 deg of the bearing arc. It is also important to note that at the point where the pressure reverses from above to below atmospheric pressure, the pressure gradients remain constant.

Fig. 12 shows circumferential pressure distribution at the middle of a bearing as a function of load. This figure demonstrates the shift in positive and negative pressure with load and also the ratios of positive to negative pressure. These ratios increase with load which shows that it is quite difficult to establish high negative pressure.

Figs. 13 and 14 show the effect of clearance on pressure distribution. A comparison also can be made between runs reported in Figs. 13 and 14 where different bearings operate at the same speed and approximately the same load. The bearing reported in Fig. 14 had 64 per cent of the projected area of the bearing reported in Fig. 13.

Figs. 15 and 16 summarize the results of various tests and compare theoretical with experimental results. Fig. 15 shows Sommerfeld number versus eccentricity ratio and minimum film thickness. Note that the load-carrying capacity of compressible bearings is higher than that of incompressible bearings.

It should be noted that the theoretical compressible curve, Fig. 15, would be shifted further downwards if energy considerations were included in the analysis. In this analysis, viscosity was considered to remain constant.

Good correlation appears to exist up to an eccentricity ratio of about 0.7. The discrepancies at higher eccentricity ratios are contributed to mostly by measurement error, manufacturing tolerances, misalignment, and the assumption of the thermodynamic path, all of which produce large effects on thin films.

AXIAL IN PLANE OF LOAD

CIRCUMFERENTIAL AT VARIOUS POSITIONS ALONG LENGTH OF BEARING
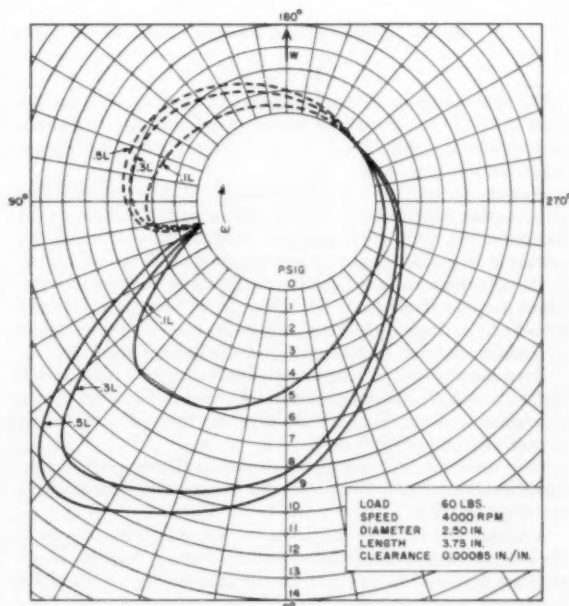
Fig. 11 Experimental pressure distribution
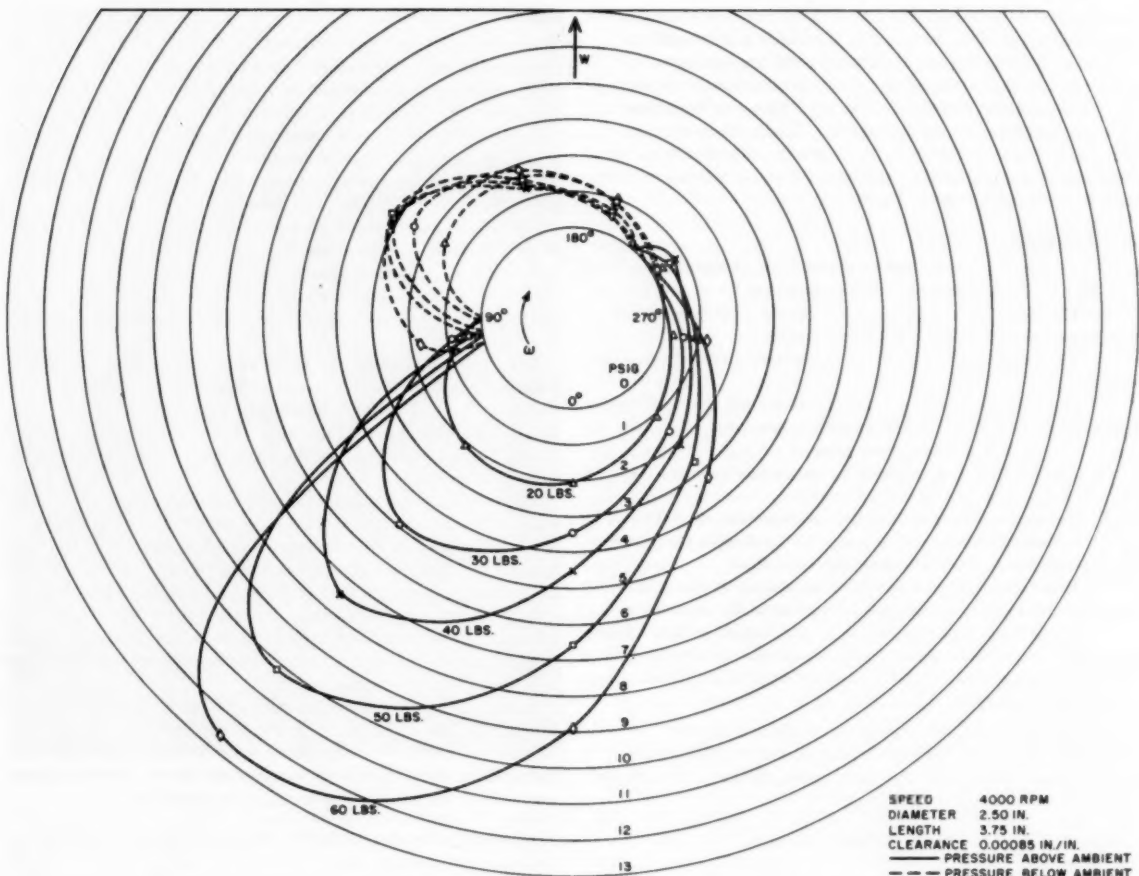
Fig. 12 Circumferential pressure distribution at middle of bearing as a function of load obtained experimentally

The measurements of film thickness and attitude angle are the most difficult of all measurements in bearing investigations, particularly at small values of film thickness where calibration errors become very large.

Manufacturing errors and tolerances, minute scratches, and other surface imperfections become especially influential when the minimum film thickness approaches their magnitude. Since the surface is irregular, expansions and contractions of the fluid occur in the asperities, causing pressure distributions considerably different from the theoretical. The discrepancy between the tests of 2-in-diam bearings of different clearances, Figs. 15 and 16, tends to confirm the belief that surface irregularities are important.

Even the slightest misalignment of the bearing, or deflection of the shaft, will produce an imperfect lubricant path, resulting in a divergence from the theory, which says that the shaft and bearing are perfectly aligned cylinders.

As to the assumed thermodynamic path in this analysis, further experiments are under way to test this assumption, since it affects the whole analysis.

Although the correlation between theory and performance of the 2-in. bearing with 0.00095 clearance (Fig. 15) is poor, the results are presented to illustrate the cumulative effect of the aforementioned imperfections.

Based on these divergences of practice from theory, it is to be expected that experiment will diverge from analysis when conditions become extreme, such as at high eccentricities.

An estimation of the load capacity of a journal bearing lubricated with air can be obtained from the results shown in Fig. 15. A safe assumption is that the Sommerfeld number must not be smaller than 0.045. This value would give an eccentricity ratio of about 0.8.

Further, assuming that clearance equals 0.001 in/in. and air viscosity equals 2.7×10^{-9} lb-sec/in.^{3}, one obtains then a maximum load of $P_{max} = 0.06 N$. N is in rps and P_{max} in psi. When $N = 60$ rps, P_{max} is about 3.6 psi. For $N = 167$ rps, P_{max} would be about 10 psi.

Fig. 16 shows eccentricity ratio versus attitude angle for the configurations tested. For comparison, theoretical loci are also drawn for compressible and incompressible fluids.

At low eccentricity ratios (approximately 0.25 or lower) the bearings exhibited fluid instability whose frequency was one half or very nearly one half the rotational frequency. This instability can be eliminated by grooving the bearing or the shaft (5). The grooving provides hydrodynamic restoring forces; however, it generally reduces the load-carrying capacity.

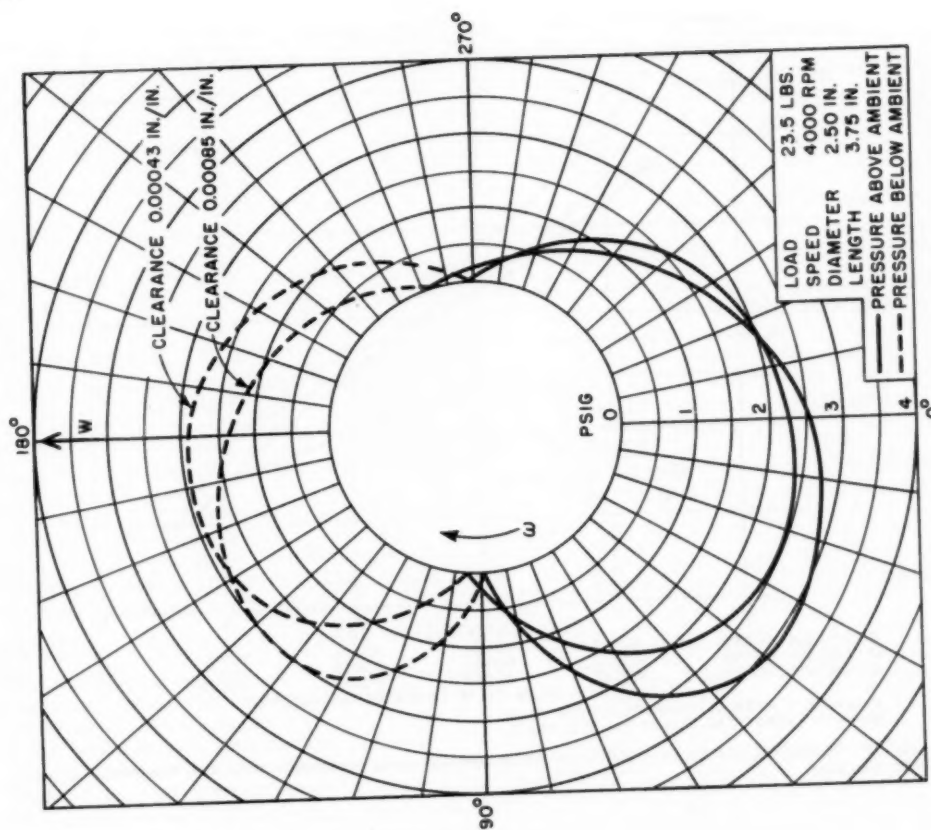Fig. 14 Experimental circumferential pressure distribution at middle of bearing as affected by clearance
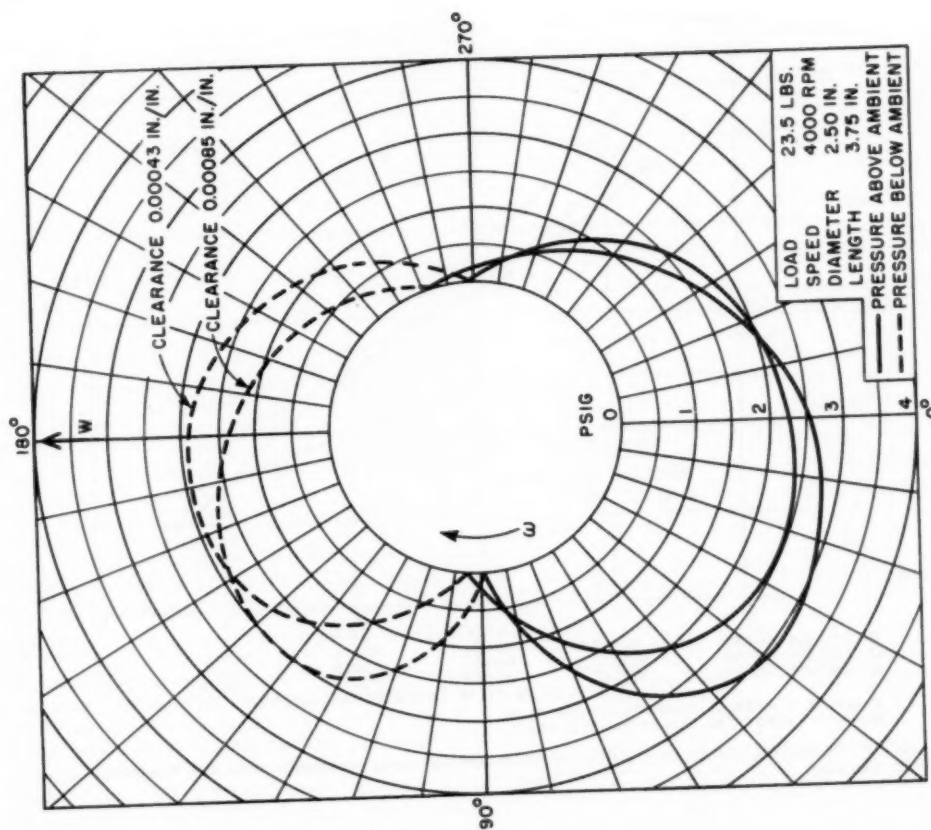

Fig. 13 Experimental circumferential pressure distribution at middle of bearing as affected by clearance
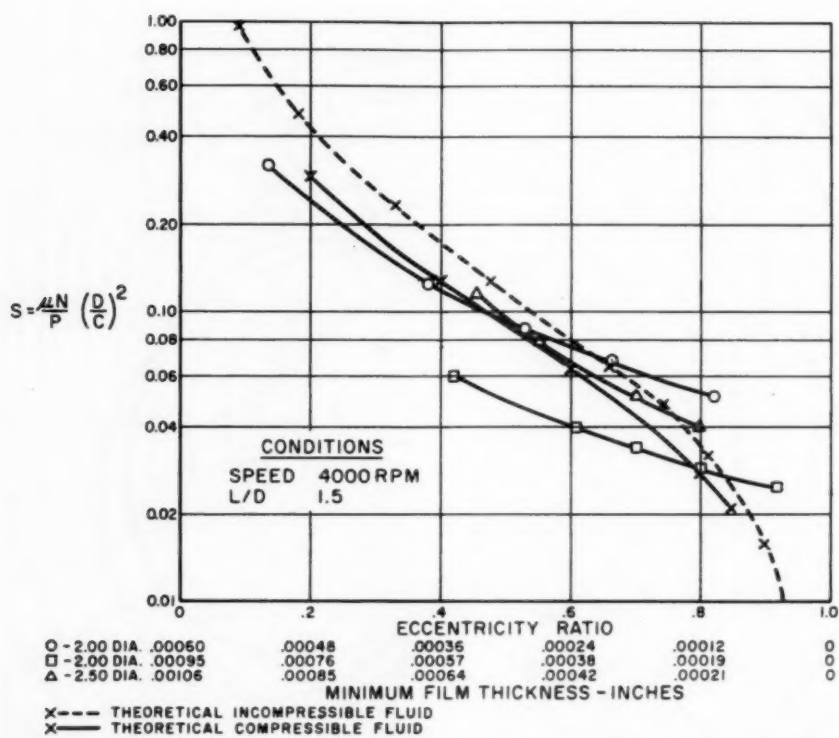
Fig. 15 Theoretical and experimental Sommerfeld number versus eccentricity ratio and minimum film thickness
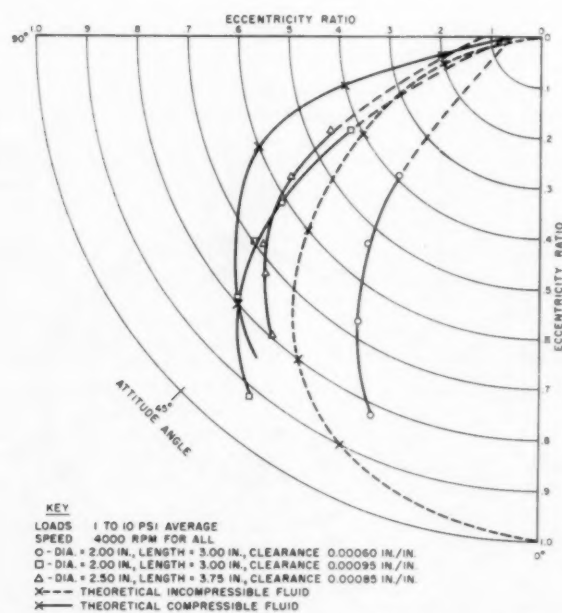

Fig. 16 Theoretical and experimental eccentricity ratio versus attitude angle

Conclusions

1 Compressibility should be included in the analysis of air lubricated hydrodynamic bearings if close accuracy is required.

2 The turbulent incompressible case is closer than the laminar incompressible case to the compressible solution.

3 Solutions for finite-width journal bearing may be obtained conveniently by numerical methods using high-speed computers.

4 There is good correlation between theoretical and experimental results. Compare Fig. 4 with Figs. 11 and 12, and the summary of results in Figs. 15 and 16.

5 Appreciable load can be carried by hydrodynamic compressible-fluid bearings. The following equation can be used for obtaining order of magnitude

$$P_{max} = 0.06 \, N$$

where

$$P_{max} = \text{psi}, \quad N = \text{rps}$$

For a given load using the foregoing equation, one can find the size of the bearing readily.

6 For any specific load it is more desirable to go to a larger diameter bearing than to a longer bearing.

7 In order to eliminate instability at light loads, grooving may be incorporated. Unless the grooves are designed properly, the load-carrying capacity tends to be reduced (5).

Recommendations

1 A study should be made on methods of starting and stopping hydrodynamic air bearings. This study should include hydrostatic starting devices, fluid injection, and so on.

2 Analyses should be undertaken to study the effect of hydrodynamics on hydrostatics. This work has been initiated by the authors.

3 Hydrodynamic journal bearings have a very low spring constant and virtually no damping constant. A study should be made to determine these constants as a function of "squeeze film" velocity.

4 A study also should be made to determine the load-carrying capacity of misaligned hydrodynamic gas journal bearings.

Acknowledgments

The authors wish to express their appreciation to Mrs. P. Mall who programmed the analysis on the IBM 650 digital computer, to Mr. L. Licht who was primarily responsible for the design of the test equipment, and to Messrs. J. V. Foley and H. E. Marx who conducted the tests.

Bibliography

1 "Experiments With an Air-Lubricated Journal," by A. Kingsbury, *Journal of the American Society of Naval Engineers,* vol. 9, 1897, pp. 267–292.

2 "Some Laboratory Applications of the Low Friction Properties of the Dry Hypodermic Syringe," by H. F. Bruback, *Review of Scientific Instruments,* vol. 18, May 1947, pp. 363–366.

3 "Gleitlager mit Luftschmierung," by H. Drescher, *Zeitschrift VDI,* vol. 95, 1953, pp. 1162–1190.

4 "Low-Friction Properties of Air-Lubricated Bearings," by D. D. Fuller, Annals of the New York Academy of Science, series 2, vol. 15, 1953, pp. 93–99.

5 "Investigation of Translatory Fluid Whirl of Vertical Shafts," by B. Sternlicht, Library of Congress, 1954.

6 "Experiments on Gas-Lubricated Journal Bearings," by M. Wildmann, ASME Paper No. 56—LUB-8.

7 "The Hydrodynamic Theory of Lubrication With Specific Reference to Air as Lubricant," by W. J. Harrison, *Transactions of the Cambridge Philosophical Society,* vol. 22, no. 3, 1913, p. 39.

8 "The Fluid Dynamic Theory of Gas-Lubricated Bearings," by J. S. Ausman, TRANS. ASME, vol. 79, 1957, pp. 1218–1224.

9 "Journal Bearing Operation at Super-Laminar Speeds," by M. I. Smith and D. D. Fuller, TRANS. ASME, vol. 78, 1956, p. 469.

Discussion

J. S. Ausman.[4] The authors have made a worthwhile contribution to the science of hydrodynamic gas lubrication in the form of experimental data and numerical solutions to the theoretical equations. It should be pointed out, however, that the results presented, both experimental and theoretical, are confined to relatively small values of the parameter $\lambda = \dfrac{6\mu\omega D^2}{\gamma P_0 C^2}$ as used by this writer (1)[5] or $G = \dfrac{6\mu\omega D^2}{P_0 C^2}$ as used by Ford, Harris, and Pantall (2). For this reason, some of the conclusions reached by the authors must be used cautiously because, while they apply in the region $0 < \lambda < 1$, they do not necessarily apply for higher values of λ. For example, the conclusion that load is directly proportional to speed, while true for these small values of λ, is not true for high values of λ. In fact, for $\lambda > 10$ load is nearly independent of speed (1, 2, 3).
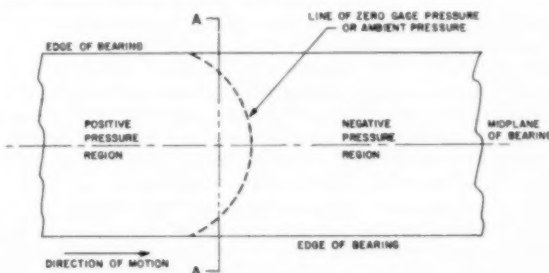
Fig. 17 "Unrolled" journal bearing showing shape of zero (gage) pressure line
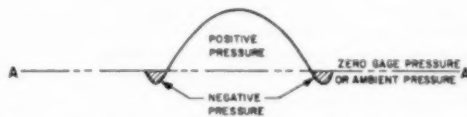
Fig. 18 Axial pressure distribution along line A-A in Fig. 17

Caution is also advised in interpreting the axial pressure distributions. Figs. 4 and 11 imply that points of zero (gage) pressure form a straight line along the length of the bearing. Such is not always the case as has been shown experimentally by Wildmann (3) and verified theoretically by this writer (4). In general, the line of zero pressure in gas-lubricated journal bearings is a curve convex in the direction of motion with the outer ends lagging behind the central portion of the curve (see Fig. 17). If pressure measurements are made along the line A-A in Fig. 17, the peculiar pressure distribution of Fig. 18 is observed. Pressure distributions of exactly this form are reported by Wildmann (3) and were substantiated by V. McNeilly of du Pont in the oral discussion of Wildmann's paper at the 1956 Lubrication Conference in Atlantic City. For the small λ values which the authors considered, the amount of lag is only about 10 deg. This is small compared to the 12 × 12 grid mesh width of 30 deg, and it may also be small with respect to the angular spacing between experimental pressure measurements. If so, this could account for the authors' failure to observe or at least their reluctance to report this phenomenon.

Bibliography

1 "The Fluid Dynamic Theory of Gas-Lubricated Bearings," by J. S. Ausman, Trans. ASME, vol. 79, 1957, p. 1218.

2 "Principles and Applications of Hydrodynamic-Type Gas Bearings," by G. W. K. Ford, D. M. Harris, and D. Pantall, paper presented at the IME General Meeting in London, October 26, 1956.

3 "Experiments on Gas-Lubricated Journal Bearings," by M. Wildmann, ASME Paper No. 56—LUB-8, presented at the Joint ASME-ASLE Lubrication Conference, October, 1956.

4 "The Finite Gas-Lubricated Journal Bearing," by J. S. Ausman, paper presented at the IME Conference on Lubrication and Wear, London, October, 1957.

G. K. Fischer[6] and J. Cherubim.[7]   It was indeed a pleasure to read this paper in so far as it substantiates to a large extent the work done in the past couple of years by us using the analyses of Karelitz, Needs, and others.

Our tests were run on $1/2$-in. and $3/4$-in. diameter shafts at speeds ranging from 100,000 to 165,000 rpm with various amounts of hydrostatic and hydrodynamic elements. Shaft loads ranged up to 15 lb. In order to be able to operate at these speeds and these loads, prediction and control of the instability called "half-speed whirl" were required. Based on extension of existing theories, we have had a great deal of success in being able to predict, and control the phenomenon of half-speed whirl under hydrodynamic as well as hydrodynamic-hydrostatic operation. Other instabilities of importance have also been encountered which required solutions. The problems of eliminating these other instabilities have been solved to a large extent, some solutions being rigorously mathematical, others being empirical.

Recognizing the limitations of high speeds in regard to instabilities in full journal bearings, we started our program with grooved designs as the authors have suggested as a means for the elimination of the instability at small eccentricity ratios. It is generally acknowledged that grooving reduces the load carrying capacity. However, we have found that the Sommerfeld number at eccentricities of about 0.8 for the design we are using is substantially the same as that reported in this paper for full journal bearings. It should be stressed for clarity that the equation cited in the conclusions of the paper ($P = 0.06N$) should be used only for clearance ratios of 0.001 in/in.

We have found sufficiently good correlation using incompressible theory and the charts of Needs and Karelitz for the prediction of hydrodynamic load carrying capacity of a practical design on the conservative side. As the authors point out in their paper, programming and computer time can be relatively expensive. We consider the sacrifice in accuracy by not using compressible theory justifiable on the basis of less time and expense for the present. However, we are looking forward to the paper that the authors intend to publish concerning the programming of their calculations on the IBM 650 digital computer.

The recommendations by the authors encompass the main items of our investigations at Stratos. Since hydrostatic pressure is available to some extent in our intended applications, our investigations along these lines have shown the feasibility and predictability in regard to starting and stopping, as well as augmenting of hydrodynamic load carrying capacity. Unfortunately, the results obtained are not directly comparable to the results presented in the authors' paper but will be part of the subject of a paper that we intend to offer for publication. The attitude-eccentricity curves that we have measured for our grooved bearing with hydrostatic pressure require more discussion than we could offer here. However, an actual cathode-ray oscilloscope photograph of the attitude-eccentricity curve, which is typical of what

[6] Research Engineer, Stratos Division, Fairchild Engine & Aircraft Corporation, Bayshore, N. Y. Assoc. Mem. ASME.

[7] Research Engineer, Stratos Division, Fairchild Engine & Aircraft Corporation, Bayshore, N. Y.
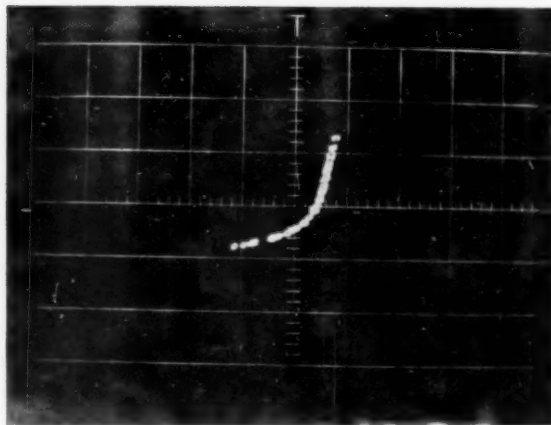
Fig. 19
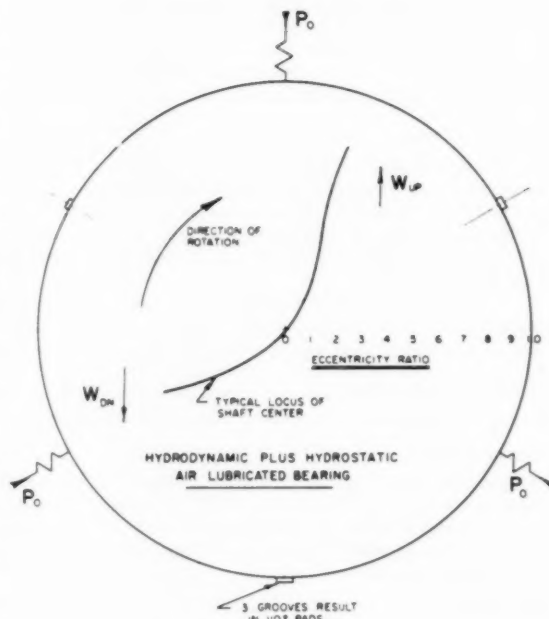
Fig. 20

we have been measuring over the past year, is shown in Fig. 19. The sketch, Fig. 20, describes the photograph and geometry of the bearing. It includes instrument calibration corrections.

Concerning damping, it may be helpful to the authors to know that with hydrostatic air pressure of 20 psig on a grooved bearing, $1/2$ in. in diameter, with a shaft weight of 6 oz and a diametral clearance of 0.001 in., we have measured damping coefficients in the order of 0.1 sec/in. A typical trace (Fig. 21) is an actual photograph from our C.R.O.

In regard to the capacitance probe used for the determination of film thickness, it is not quite clear as to how the results presented took into account the apparent interaction between the bearing surfaces and the probe surface. If the probe was placed within the reported bearing length, then an apparent reduction in load carrying capacity should have taken place due to the
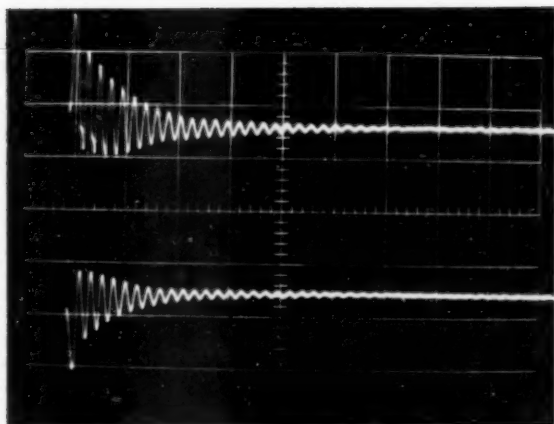
Fig. 21

added clearance between journal and probe. However, if the probe was located outside of the reported length, it could have acted as another hydrodynamic bearing and caused an apparent increase in load carrying capacity. This effect is a function of the area presented to the journal and was probably kept to a minimum by the author. The size of the probe area would also affect the accuracy to which the attitude-eccentricity curves would be plotted. In our active work of adapting the Fielden Proximity Meter to high speed micro-displacement phenomena we found a great importance in regard to location and design of the capacitance probe.

L. M. Tichvinsky.[8] It is very possible that in the future some bearings will perform with other than air gases, which may be real or ideal, monatomic, diatomic, or polyatomic. Consequently, the paper correctly indicates this by mentioning gas-lubricated journal bearings; however, the paper presents numerical solutions for finite-width journal bearings and results of an experiment conducted with air-lubricated hydrodynamic journal bearings.

Then, if real gases are used as lubricants it will be necessary to introduce some additional terms—thermodynamic ones—and refer to the law of corresponding states and the compressibility factor $Z = pv/RT$, where p is pressure, v, volume, R, gas constant, and T, absolute temperature. These will in turn determine the viscosity of the gas used.

M. Wildmann.[9] The authors should be congratulated for adding some theoretical and experimental information to the field of hydrodynamic gas lubrication. The experimental data are particularly interesting, as they cover the region of high eccentricity ratio, where practically no data were previously available. The method of obtaining numerical solutions to the gas bearing equations should be useful if extended to other gas bearing configurations, such as gas-lubricated slider bearings of finite width. Fortunately, enough theoretical work on gas-lubricated journal bearings is now available, so that the performance of these bearings can be predicted with fair accuracy without having to resort to the tedious numerical methods outlined (1 and 2).[10] This available literature enables us to explain some of the authors' statements—and question others.

[8] Professor of Mechanical Engineering, University of California, Berkeley, Calif. Mem. ASME.
[9] Research Engineer, Autonetics Division, North American Aviation, Inc., Downey, Calif.
[10] Numbers in parentheses refer to the References at the end of the discussion.

Throughout their paper, the authors compare gas-lubricated journal bearings to journal bearings using noncompressible lubricants with oil film breakdown in the region of negative pressure. This gives fair agreement with the compressible case in some particular instances, because in a gas-lubricated journal bearing, the negative pressure is always smaller than the positive pressure, thus approximating the condition of film breakdown. Actually, no lubricating film breakdown occurs in a gas-lubricated bearing, but the smaller negative pressures can be predicted from the theoretical equation. The limiting pressure distribution in a gas-lubricated journal bearing is $P_0/(1 + \epsilon \cos \theta)$. Thus, for an eccentricity of one, the maximum negative pressure is $P_0/2$, while the maximum positive pressure approaches infinity.

In Figs. 5 and 6, the authors give curves of Sommerfeld number versus eccentricity, and attitude angle versus eccentricity, for two different L/D ratios. But it is not stated for what combination of speed, viscosity, ambient pressure, and clearance ratio these curves were obtained. These curves obviously do not cover all possible combinations of these constants. For example, in the case of extremely high speeds or small ambient pressures, the attitude angle of these bearings approaches zero. Great care should be taken therefore if these curves are used for gas bearing design, and no general statements about gas bearings can be made based on these curves alone. This applies particularly to the authors' statement about relative attitude angles for the compressible and noncompressible case, and their relative load capacities. These statements are correct for the bearings investigated, but not for gas bearings in general.

In Figs. 15 and 16, the authors compare their experimental data to theoretical results. All the experimental bearings have the same L/D ratio. Nevertheless, no single theoretical curve can fit all three bearings, as each bearing has a different clearance ratio. Very good agreement between theory and experiment therefore cannot be expected.

In the conclusions, the authors give an approximate equation to determine maximum pressures. Again, this equation is correct only for the bearings analyzed, but cannot be applied to gas bearings in general. This equation implies that the maximum pressure in a gas bearing is a linear function of speed, while actually for high speeds the maximum pressure in a gas bearing approaches some finite value which is independent of speed and directly proportional to ambient pressure. Conclusion 6 also calls for some further explanation from the authors, as it would be very surprising if a bearing with a small L/D ratio is preferable to a bearing with a large L/D, except in some very limited cases.

References

1 "The Finite Gas-Lubricated Journal Bearing," by J. S. Ausman, presented at the IME Conference on Lubrication and Wear, London, October, 1957.
2 "Theory of Lubrication by Compressible Fluid With Special Reference to Air Bearing," by U. Katto and N. Soda, Proceedings of the Second Japan National Congress for Applied Mechanics, 1952, pp. 267–270.

Authors' Closure

The authors are pleased to see the extent of interest in gas-lubricated journal bearings and wish to thank the discussers for their questions and stimulating comments.

Mr. Ausman is perfectly correct in pointing out that the investigation is confined to relatively small values of the parameter λ and that the conclusions should be used with caution at high values of λ. However, our purpose in this paper was to develop numerical methods for solving the theoretical equations which we hope others will use when interested in higher values of λ. The method of solution is perfectly general and applies also to partial and grooved journal bearings. Our interest was primarily con-

cerned with high loads and sea-level atmospheric conditions.

Mr. Ausman's comments on axial and circumferential pressure distributions are also quite true. In our experiments, the indexing method used in taking circumferential pressure distributions was not sensitive enough to measure accurately this characteristic since the smallest index angle was 10 deg. In the theoretical analysis, the circumferential grid would have to be divided to approximately 120 meshes in order to pick this up. This would result in a tenfold increase in computer time with negligible gain in load accuracy. It should also be pointed out that this condition exists in incompressible hydrodynamic journal bearings.

Messrs. Fisher and Cherubim are thanked for the generous use of their own test results in discussing our paper. It is most gratifying to see that our work is substantiated by the discussers. A recent paper presented in England by Messrs. J. A. Cole and J. Kerr entitled "Observations on the Performance of Air-Lubricated Bearings" also confirms our results.

There is no direct comparison between our Figs. 6 and 16, and Messrs. Fisher and Cherubim's Figs. 19 and 20. Our paper deals with purely hydrodynamic bearings while Figs. 19 and 20 are for a special configuration of hydrodynamic-hydrostatic bearing. The reader is cautioned not to confuse them.

We have specifically presented theoretical results for both the compressible and incompressible fluid-lubricated journal bearing in Figs. 5 and 15 for comparison. These figures verify the comments made by the discussers that incompressible theory gives conservative load carrying capacity.

The damping coefficient reported by the discussers is extremely low, which points out that they were operating at low eccentricity ratios and in a region where the bearing is susceptible to instability. We have measured damping coefficients in hydrostatic gas bearings under similar supply pressures which were greater a hundredfold. It should be remembered also that the damping coefficients of gas bearings are many times smaller than incompressible fluid bearings, which makes them inherently less stable.

With regard to the question raised about the capacitance probe, it should be pointed out that our gage area was 0.03 sq in., which is $1/2$ of 1 per cent of the area of our smallest bearing. Its effect on bearing performance we feel is negligible. In the case of the $1/2$-in. bearings reported by the discussers, our probe would take up 8 per cent of the projected area, assuming an L/D of 1.5. This would introduce appreciable errors in the measurement of attitude angle and eccentricity ratio, since the width would be finite and represent an appreciable percentage of the shaft diameter.

Professor Tichvinsky is correct in pointing out that in the fu-

ture other gases than air will be used in gas bearings. The future is today and several gases have already been used. Some of these have been reported in papers by Dr. Ausman and Mr. Wildmann. It is also true that thermodynamic relationships have to be used relating pressure, density, and viscosity in the analysis.

Mr. Wildmann is quite correct in stating that the numerical solutions are very tedious and that closed solutions should be used wherever possible. However, once the numerical solutions are programmed on a high speed computer, they become easy to use. In many cases, they are the only solutions that are at the present moment available. They apply to partial and grooved bearings and also at high eccentricity ratios where closed solutions are not available.

Extensive literature has been published on incompressible hydrodynamics. On the other hand, very limited information is available on compressible hydrodynamics. We have therefore tried to compare to a limited extent these two types of fluids so that the reader can see the differences in their characteristics in journal bearings.

Figs. 5 and 6 are restricted to air as a fluid at atmospheric conditions, but in other respects are quite general. They do not include variable viscosity and density with temperature. The main purpose for plotting them in dimensionless form is to extend their range of applicability and we feel that Mr. Wildmann may have missed this point.

The Sommerfeld number has the term $(D/C)^2$ in it and therefore the comparison between theory and experiments is valid. This comparison is made in Figs. 15 and 16. The agreement between theory and experiment is not as good as we would like it to be, of course, but it does justify the analytical method used.

As for the rule of thumb equation presented in the conclusions, we do not seriously believe that anyone would design a bearing system on such limited information. The assumptions used in its derivation are given beneath Fig. 12 and include the clearance used. It is very interesting to note that Mr. Cole and Mr. Kerr in their paper quite independently come up with exactly the same rule of thumb.

Our conclusion number 6 is based upon observation of Fig. 5. It is apparent that S increases linearly with length and as a cubic function of D. There is of course a limit to this, for an infinitely short bearing cannot generate any hydrodynamic pressure and therefore carries zero load. Thus $L/D = 1/2$ is not uncommon in gas bearings.

Again, we should like to express our appreciation and thanks to those who have contributed their time and effort in the discussion of this paper.

The Variable Fluid-Property Problem in Free Convection[1]

By E. M. SPARROW[2] AND J. L. GREGG,[3] CLEVELAND, OHIO

An analysis is made for the variable fluid-property problem for laminar free convection on an isothermal vertical flat plate. For a number of specific cases, solutions of the boundary-layer equations appropriate to the variable-property situation were carried out for gases and for liquid mercury. Utilizing these findings, a simple and accurate shorthand procedure is presented for calculating free-convection heat transfer under variable-property conditions. This calculation method is well established in the heat-transfer field. It involves the use of results which have been derived for constant-property fluids, and of a set of rules (called reference temperatures) for extending these constant-property results to variable-property situations. For gases, the constant-property heat-transfer results are generalized to the variable property situation by replacing β (expansion coefficient) by $1/T_\infty$ and evaluating the other properties at $T_r = T_w - 0.38(T_w - T_\infty)$. For liquid mercury, the generalization may be accomplished by evaluating all the properties (including β) at this same T_r. It is worth while noting that, for these fluids, the film temperature (with $\beta = 1/T_\infty$ for gases) appears to serve as an adequate reference temperature for most applications. Results are also presented for boundary-layer thickness and velocity parameters.

Nomenclature

The following nomenclature is used in the paper:

a_n = coefficients in the polynomial representations of fluid properties of liquid mercury

A_1, A_2 = dimensional constants in Sutherland's formulas of Table 1, $A_1 = 362$ F abs, $A_2 = 198.7$ F abs

b = plate width

c = dimensional constant defined by Equation [6a]

c_p = specific heat at constant pressure

c_f = coefficient of friction defined by Equation [24]

F = dimensionless dependent variable defined by Equation [6b]

g = acceleration due to gravity

Gr_x = Grashof number based on x, dimensionless, see Equations [13], [13a], [13b]

Gr_L = Grashof number based on L

h = local heat-transfer coefficient, $q/(T_w - T_\infty)$

$\bar{h}$ = average heat-transfer coefficient, $Q/Lb(T_w - T_\infty)$

k = thermal conductivity

L = plate height

Nu_x = local Nusselt number, hx/k, dimensionless

$\overline{Nu_L}$ = average Nusselt number, $\bar{h}L/k$, dimensionless

p = static pressure

Pr = Prandtl number, $c_p\mu/k$, dimensionless

q = local heat-transfer rate per unit area of plate

Q = over-all heat-transfer rate, $b\int_0^L q\,dx$

T = absolute temperature

t = Fahrenheit temperature

T_f = film temperature, $(T_w + T_\infty)/2$; or $t_f = (t_w + t_\infty)/2$

u = velocity component in x-direction

u_{max} = maximum value of u across boundary layer

v = velocity component in y-direction

W = rate of fluid flow generated by free convection,

$$b\int_0^\infty \rho u\,dy$$

x = co-ordinate measuring distance along plate from leading edge

y = co-ordinate measuring distance normal to plate

y_u = distance from plate at which u_{max} occurs

β = coefficient of thermal expansion, $-\dfrac{1}{\rho}\left(\dfrac{\partial\rho}{\partial t}\right)_p$

δ_i = boundary-layer thickness defined as distance from plate at which $(T - T_\infty)/(T_w - T_\infty) = i$

η = dimensionless similarity variable defined by Equation [6a]

θ = dimensionless temperature variable $(T - T_\infty)/(T_w - T_\infty)$

μ = absolute viscosity

ν = kinematic viscosity, μ/ρ

ρ = density

τ_w = shear stress at plate surface

ψ = stream function

Subscripts

w = wall conditions

∞ = ambient conditions

r = conditions at reference temperature

Introduction

The presence of a buoyancy force is a requirement for the existence of a free-convection flow. Ordinarily, the buoyancy arises from density differences which are a consequence of temperature gradients within the fluid. Any analytical treatment of free convection must include density variations at least to the extent that a buoyancy force enters the problem.

A characteristic common to previous analytical studies of free convection has been the neglect of all fluid-property variations, except for the essential density differences just noted.[4] Such a simplified treatment does not appear unreasonable when the temperature differences involved are small. This intuitive feeling has been corroborated in a formal manner by Ostrach (2).[5] For situations where there are large temperature differences, the

---

[4] The fluid properties entering the problem are the thermal conductivity, viscosity, specific heat, and density.

[5] Numbers in parentheses refer to the Bibliography at the end of the paper.

adequacy of the results derived from the constant-property analysis has been in doubt.

An analytical treatment of the variable-property problem, including numerical solutions, first appears to have been given in the thesis (1) from which this paper is taken. Some time after the appearance of reference (1), a much less extensive study was described in a Russian article by Tanaev (3). There is no significant overlap between reference (3) and the presentation here.

The analysis is made for an isothermal vertical plate, and the flow is taken to be laminar. For a large number of specific cases, numerical solutions of the boundary-layer equations appropriate to the variable-property situation were carried out for gases. It was initially planned to also study the variable-property problem for several liquid metals. However, the nature of the property variations and the exceedingly time-consuming numerical computations forced the current study to be limited to one liquid metal; namely, mercury.

Prime attention is focused on the heat transfer. As has already been described in the Abstract, our approach was to generalize the findings for specific cases into a simple and accurate shorthand procedure for computing engineering results under variable-property conditions. The generalization was made not only for the heat transfer, but also for several velocity and boundary-layer thickness parameters. Those who are interested primarily in results are invited to pass over the section on analysis.

Analysis

Physical Model and Co-Ordinates. The physical model and the co-ordinate system are displayed in an elevation view in Fig. 1. Two physical situations are shown which come within the scope of the analysis. The left-hand sketch depicts the case where the wall temperature T_w exceeds the ambient temperature T_∞. Under these circumstances the free-convection motion is upward as shown. The right-hand sketch shows the situation where T_w is lower than the ambient temperature T_∞; and the fluid flow is downward along the plate.
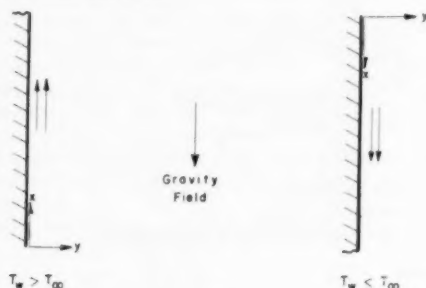
Fig. 1   Co-ordinate systems

If the co-ordinate systems are taken as indicated, the mathematical distinction between the two situations vanishes when the conservation equations, as written later, are made dimensionless. Hence separate analyses need not be made. Since it seems easier to visualize occurrences associated with the hot-wall case, i.e., $T_w > T_\infty$, the analysis will be directed toward that situation. However, the results will be presented in a manner applicable to both $T_w > T_\infty$ and $T_w < T_\infty$.

Conservation Laws. The equations expressing conservation of mass, momentum, and energy for steady flow in a boundary layer on a vertical plate are

$$\frac{\partial}{\partial x}(\rho u) + \frac{\partial}{\partial y}(\rho v) = 0 \dots \dots \dots \dots [1]$$

$$\rho\left(u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y}\right) = g(\rho_\infty - \rho) + \frac{\partial}{\partial y}\left(\mu\frac{\partial u}{\partial y}\right)\dots[2]$$

$$\rho c_p\left(u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y}\right) = \frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right)\dots\dots\dots[3]$$

Viscous dissipation and work against the gravity field have been neglected. The driving force for the free-convection motion is the buoyancy term $g(\rho_\infty - \rho)$; it is formed by combining the pressure gradient $\partial p/\partial x$ with the body force ρg.

The boundary conditions appropriate to the problem are

$$\left.\begin{array}{c} v = 0 \\ u = 0 \\ T = T_w \end{array}\right\} y = 0 \qquad \left.\begin{array}{c} u = 0 \\ T = T_\infty \end{array}\right\} y = \infty \dots\dots[4]$$

where T_w and T_∞ are prescribed constants.

The customary approach followed in the constant-property analysis is first to replace the density difference by a temperature difference, and then to assume that ρ, μ, c_p, and k are constant.

The present analysis continues without placing any restrictions on the nature of the property variations. The solution of Equation [1], as usual, may be written in terms of a stream function ψ defined by the relations

$$\frac{\rho}{\rho_w}u = \frac{\partial\psi}{\partial y}, \quad \frac{\rho}{\rho_w}v = -\frac{\partial\psi}{\partial x}\dots\dots\dots\dots[5]$$

where ρ_w, the fluid density at the wall, is regarded as a constant. Then the velocity components u and v which appear in Equations [2] and [3] are replaced in favor of the stream function ψ. The result of the substitution is a rather complicated looking pair of simultaneous partial differential equations for ψ and T as functions of x and y. Rather than deal with these two formidable partial differential equations directly, experience suggests a method of transforming them to ordinary differential equations, which are easier to solve. In the usual terminology of boundary-layer theory, such a transformation is called a similarity transformation.

Reduction to Ordinary Differential Equations. A new independent variable η, called a similarity variable, is defined by

$$\eta = cx^{-1/4}\int_0^y \frac{\rho}{\rho_w}\,dy, \quad c = \left[\frac{g(\rho_\infty - \rho_w)/\rho_w}{4\nu_w^2}\right]^{1/4}\dots[6a]$$

New dependent variables F and θ are given by

$$F(\eta) = \left(\frac{\psi}{x^{1/4}}\right)\left(\frac{1}{4\nu_w c}\right), \quad \theta(\eta) = \frac{T - T_\infty}{T_w - T_\infty}\dots\dots[6b]$$

The function θ is a dimensionless temperature and F is related to the velocities in the following way

$$u = 4\nu_w c^2 x^{1/2}F', \quad v = \left(\frac{\rho_w}{\rho}\right)\left(\frac{\nu_w c}{x^{1/4}}\right)[\eta F' - 3F]\dots[7]$$

The primes represent differentiation with respect to η.

Under the transformation Equations [6a] and [6b], the partial differential equations for ψ and T become

$$\left.\begin{array}{c} \dfrac{d}{d\eta}\left[\dfrac{\rho\mu}{\rho_w\mu_w}F''\right] + 3FF'' - 2(F')^2 + \dfrac{(\rho_\infty/\rho) - 1}{(\rho_\infty/\rho_w) - 1} = 0 \\[2mm] \dfrac{d}{d\eta}\left[\dfrac{\rho k}{\rho_w k_w}\theta'\right] + 3\mathrm{Pr}_w\left(\dfrac{c_p}{c_{pw}}\right)F\theta' = 0 \end{array}\right\}\dots[8]$$

The Prandtl number is represented by Pr and the subscript w denotes conditions at the wall $(y = 0)$. The boundary conditions, Equation [4], transform to

$$F = 0 \atop \left.\begin{matrix} F = 0 \\ F' = 0 \\ \theta = 1 \end{matrix}\right\} \eta = 0 \qquad \left.\begin{matrix} F' = 0 \\ \theta = 0 \end{matrix}\right\} \eta = \infty \quad \dots\dots [9]$$

It may be noted that the transformation relations, as well as the resulting ordinary differential equations, can be reduced directly to the well-known equations for the constant-property fluid by making the usual approximations.

Simplifications Associated With $p = \rho RT$. For the special case of fluids which obey the perfect gas law, $p = \rho RT$, there are certain simplifications which can be introduced into the analysis. Consider first the buoyancy force $g(\rho_\infty - \rho)$ which appears in Equation [2]. For a perfect gas, this density difference can be transformed into a temperature difference without making any approximations, simply by noticing that both ρ_∞ and ρ are to be evaluated at the same static pressure (same x). With this in mind, it follows directly that

$$\rho_\infty - \rho = \frac{\rho}{T_\infty}(T - T_\infty) \dots\dots\dots [10]$$

In the constant-property analysis, it has been customary to write

$$\rho_\infty - \rho = \beta\rho(T - T_\infty)$$

where β is the coefficient of expansion. So, it is seen that the buoyancy force used in the constant-property analysis is precisely correct for a perfect gas, provided that β is replaced by $1/T_\infty$.

Further, in the differential Equations [8], the dimensionless buoyancy force (the last term) can be simplified to

$$\frac{(\rho_\infty/\rho) - 1}{(\rho_\infty/\rho_w) - 1} = \frac{T - T_\infty}{T_w - T_\infty} = \theta \dots\dots\dots [10a]$$

So, Equations [8] become

$$\left.\begin{matrix} \dfrac{d}{d\eta}\left[\dfrac{\rho\mu}{\rho_w\mu_w}F''\right] + 3FF'' - 2(F')^2 + \theta = 0 \\[2mm] \dfrac{d}{d\eta}\left[\dfrac{\rho k}{\rho_w k_w}\theta'\right] + 3\mathrm{Pr}_w\left[\dfrac{c_p}{c_{pw}}\right]F'\theta = 0 \end{matrix}\right\} \dots [8a]$$

It is worth while noting that for a very special sort of variable-property gas, Equations [8a] reduce to the equations for a constant-property fluid. Consider a gas having the property variations: $p = \rho RT$, $\rho\mu = \mathrm{const}$, $\rho k = \mathrm{const}$, $c_p = \mathrm{const}$. For such a gas

$$\rho\mu/\rho_w\mu_w = 1, \quad \rho k/\rho_w k_w = 1, \quad \mathrm{Pr} = \mathrm{const}$$

With these assumptions, Equations [8a] simplify tremendously, and there results

$$\left.\begin{matrix} F''' + 3FF'' - 2(F')^2 + \theta = 0 \\[2mm] \theta'' + 3\mathrm{Pr}\,F\theta' = 0 \end{matrix}\right\} \dots\dots [8b]$$

But, these equations are precisely those for the constant-property problem! Also identical are the boundary conditions. So, from the mathematical point of view, the constant-property problem is identical to that for the special variable-property fluid: $p = \rho RT$, $\rho k = \mathrm{const}$, $\rho\mu = \mathrm{const}$, $c_p = \mathrm{const}$. All solutions which have been obtained for the constant-property differential equations become available for this special variable-property fluid.

It is interesting to note that a similar finding applies in forced convection.

Heat-Transfer Parameters and Grashof Numbers

Local Heat Transfer. The local heat flux q from the surface to the fluid may be calculated using Fourier's law. After introducing the dimensionless variables from Equations [6a] and [6b], the expression for q becomes

$$q = -\left[k\frac{\partial T}{\partial y}\right]_{y=0} = -k_w(T_w - T_\infty)cx^{-1/4}\left[\frac{d\theta}{d\eta}\right]_{\eta=0} \dots [11]$$

The derivative $[d\theta/d\eta]_{\eta=0}$, normally abbreviated $\theta'(0)$, is found from the solutions of Equations [8]. A dimensionless representation of the results is achieved by use of the local heat-transfer coefficient and local Nusselt number, which are written in the usual way as

$$h \equiv \frac{q}{T_w - T_\infty}, \quad \mathrm{Nu}_x \equiv \frac{hx}{k} \dots\dots\dots [12]$$

Further, a generalized Grashof number which is applicable to both constant and variable-property fluids is defined by

$$\mathrm{Gr}_x \equiv \frac{gx^3\{|\rho_\infty - \rho_w|/\rho_w\}}{\nu^2} \dots\dots\dots [13]$$

The absolute magnitude sign removes the necessity for separate consideration of $T_w > T_\infty$ and $T_w < T_\infty$. This Grashof number contains the density difference (the direct driving force) rather than the usual temperature difference.

Using these definitions of h, Nu_x, and Gr_x the local heat flux given by Equation [11] becomes

$$\mathrm{Nu}_{x,w} = \frac{[-\theta'(0)]}{\sqrt{2}}\,\mathrm{Gr}_{x,w}^{1/4} \dots\dots\dots [14]$$

where the second subscript on the Nusselt and Grashof numbers indicates the location at which k and ν are evaluated.

Over-All Heat Transfer. The over-all heat transfer Q is found by integrating Equation [11]. So

$$Q = b\int_0^L q\,dx \dots\dots\dots [15]$$

where b is the plate width. Introducing the following dimensionless groups

$$\bar{h} \equiv \frac{Q}{bL(T_w - T_\infty)}, \quad \overline{\mathrm{Nu}}_L \equiv \frac{\bar{h}L}{k},$$

$$\mathrm{Gr}_L \equiv \frac{gL^3\{|\rho_\infty - \rho_w|/\rho_w\}}{\nu^2} \dots\dots [16]$$

leads to the following dimensionless result for the over-all heat transfer

$$\overline{\mathrm{Nu}}_{L,w} = \frac{4}{3}\frac{[-\theta'(0)]}{\sqrt{2}}\,\mathrm{Gr}_{L,w}^{1/4} \dots\dots\dots [17]$$

The Grashof Number. The generalized Grashof number defined by Equation [13] arises naturally from the analysis of the general variable-property fluid. The form taken by this definition for certain special cases is of interest.

For instance, for the constant-property analysis, Equation [13] simplifies to

$$\mathrm{Gr}_x = \frac{g\beta|T_w - T_\infty|x^3}{\nu^2} \dots\dots\dots [13a]$$

This is identical to usual Grashof-number definition.

For a perfect gas, the Grashof number becomes

$$\mathrm{Gr}_x = \frac{g|T_w - T_\infty|x^3}{T_\infty\nu^2} \dots\dots\dots [13b]$$

Description of Gases Studied

In previous variable-properties analyses for gases, it has been common to use idealized forms of the property variations. There have been a few instances where real gas properties were used. In the present study, both idealized and real gases have been included, the real gas being a close approximation to air. Table 1 describes the five gases to be considered here in the order (reading from left to right) in which they will be discussed. All are seen to obey the perfect gas law, $p = \rho R T$. Absolute temperatures are used exclusively throughout the analysis for gases.

$$\frac{\rho\mu}{\rho_w\mu_w} = \frac{\rho k}{\rho_w k_w} = \left[\theta\left(1 - \frac{T_\infty}{T_w}\right) + \frac{T_\infty}{T_w}\right]^{-1/4} \dots [18]$$

where we have introduced the dimensionless temperature $\theta = (T - T_\infty)/(T_w - T_\infty)$. In light of this, it may be observed that before commencing with a solution of Equations [8a] to find F and θ, it is necessary first to specify the value of T_w/T_∞ (as well as of the Prandtl number). The appearance of this temperature ratio is associated with the variable-property problem. But, the fact that only a temperature *ratio* appears is actually a considerable

Table 1 Description of gases A, B, C, D, and E

Gas A	Gas B	Gas C	Gas D	Gas E
$p = \rho R T$	$p = \rho R T$	$p = \rho R T$	$p = \rho R T$	$p = \rho R T$
$k \sim T^{1/4}$	$k \sim T^{3/4}$	$\rho k = \text{const}$	$k \sim \dfrac{T^{3/2}}{T + A_1}$	$k \sim \dfrac{T^{3/2}}{T + A_1}$
$\mu \sim T^{1/4}$	$\mu \sim T^{3/4}$	$\rho\mu = \text{const}$	$\mu \sim \dfrac{T^{3/2}}{T + A_1}$	$\mu \sim \dfrac{T^{3/2}}{T + A_2}$
$c_p = \text{const}$	$c_p = \text{const}$	$c_p = \text{const}$	$c_p = \text{const}$	$c_p \begin{matrix} = b_0 + b_1 T \\ = b_2 + b_3 T \end{matrix}$
$\text{Pr} = \text{const}$	$\text{Pr} = \text{const}$	$\text{Pr} = \text{const}$	$\text{Pr} = \text{const}$	$\text{Pr} = \text{variable}$

Gases A, B, and C represent simple idealizations of real gas behavior. The power-law variations for k and μ are commonly used approximations. The assumption of constant specific heat and Prandtl number is included because, for real gases, the variation of these properties is small compared to those of k, μ, and ρ.

The Sutherland-type formula used to describe the conductivity and viscosity variations of gas D is closer to reality than are the simple power laws. Since c_p and Pr are maintained as constants for gas D, the same Sutherland formula is used for k and μ. The constant A_1 was assigned the value 362 F abs, a number appearing in Glassman and Bonilla's (4) correlation of thermal-conductivity data for air.

Gas E is a close approximation to air. Hence variations of c_p and Pr are included, and different Sutherland formulas are used for k and $\mu(A_1 \neq A_2)$. The value 198.7 F abs (110.4 K) was selected for A_2 in accord with the NBS tables (5). The specific-heat data, also from the NBS tables, were fitted by two straight lines over the temperature range studied. The variation of the Prandtl number was not specified explicitly, since it is determined when c_p, k, and μ are given.

Heat-Transfer Results for Gases

Heat-transfer results for gases A through E will be presented for a large number of special cases. Utilizing these results, a rapid and accurate shorthand method will be presented for calculating heat transfer to gases under variable-property conditions. The calculation procedure involves the use of results which have been derived for constant-property fluids, and of a rule for extending these constant-property results to variable-property situations. This rule is commonly called a reference temperature.

The following approach will be used here in establishing a reference-temperature rule for heat transfer. First, using the numerous numerical calculations for gas A, a reference-temperature rule will be derived. Then tests of the wider applicability of his reference temperature result will be made using the less numerous solutions for gases B, C, D, and E.

Results for Gas A. In order to compute the heat transfer, it is necessary to solve the differential Equations [8a] subject to the boundary conditions [9]. The gas properties appearing in Equations [8a] may be evaluated for gas A with the aid of Table 1. So, it is seen that $c_p/c_{pw} = 1$, Pr = const, and

simplification, since for the general variable-property fluid it is necessary to specify T_w and T_∞ separately.

Numerical solutions[6] of Equations [8a] for gas A were carried out for a wide range of values of T_w/T_∞ for Pr = 0.7 and for selected values of T_w/T_∞ for Pr = 1.0. The heat-transfer results corresponding to these solutions are listed in Table 2. The Nusselt and Grashof numbers, dimensionless parameters, are given by Equations [12] and [13b], respectively.

Table 2 Heat-transfer results for gas A

T_w/T_∞	$\mathrm{Nu}_{x,w}/\mathrm{Gr}_{x,w}{}^{1/4}$ or $\overline{\mathrm{Nu}}_{L,w}\left/\dfrac{4}{3}\,\mathrm{Gr}_{L,w}{}^{1/4}\right.$	
	Pr = 0.7	Pr = 1.0
4	0.371	. . .
3	0.368	0.418
5/2	0.366	. . .
2	0.363	. . .
3/4	0.348	. . .
1/2	0.339	. . .
1/3	0.330	0.375
1/4	0.323	. . .

Now, we proceed to generalize these results. Attention is first focused on Pr = 0.7. From an analytical solution for the *constant-property fluid*, it is found that the heat transfer (for Pr = 0.7) is

$$\mathrm{Nu}_x/\mathrm{Gr}_x{}^{1/4} = 0.353 = \overline{\mathrm{Nu}}_L\left/\frac{4}{3}\,\mathrm{Gr}_L{}^{1/4}\right. \dots\dots [19]^7$$

Here, the Grashof number is given by Equation [13a]. We are immediately led to ask whether there is some way by which the constant-property result, Equation [19], can be made to coincide with the variable-property results appearing in Table 2. It may be observed that the temperature for evaluating k, ν, and β in Equation [19] is at our disposal. With regard to β, it has already been shown that it is proper to replace β by $1/T_\infty$ when perfect gases are involved. It is a matter of making a few trials to determine the temperature for evaluating k and ν of Equation [19] which gives the best agreement between the constant-property

---

[6] All numerical integrations were carried out on an IBM Card Programmed Calculator using a technique presented in detail in appendix B of reference (2).

[7] This result may also be thought of as applying to gas A when $T_w/T_\infty \rightarrow 1$.
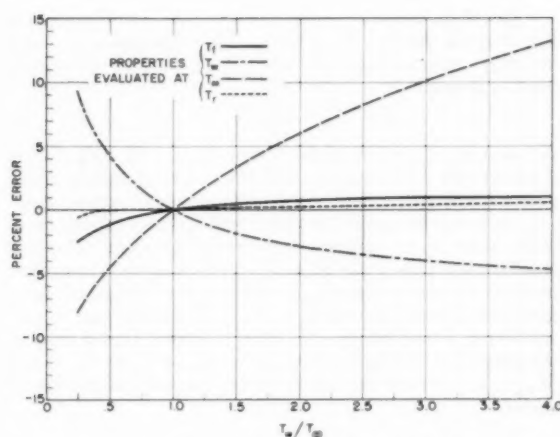
Fig. 2 Errors made in prediction of heat transfer for gas A by evaluating constant-property result at various temperatures

Fig. 3 Results of constant-property analysis for Prandtl-number range of gases

result and the variable-property findings of Table 2. This temperature, which may be termed a reference temperature T_r, is found to be

$$T_r = T_w - 0.38(T_w - T_\infty), \quad \beta = 1/T_\infty \ldots \ldots [20]$$

As shown by the dashed line in Fig. 2, the error in the heat transfer predicted from the constant-property result by using this reference temperature is at most 0.6 per cent over the entire range $1/4 \leq T_w/T_\infty \leq 4$. Fig. 2 also shows the errors in the heat transfer predicted from the constant-property result when k and ν are evaluated at the wall temperature, at the ambient temperature, and at the film temperature T_f. β is taken as $1/T_\infty$ in all cases. It is noteworthy that the use of T_f as a reference temperature provides heat-transfer results which would be adequate for almost all engineering purposes.

The results listed in Table 2 for Pr = 1 were obtained to check whether the reference temperatures found for Pr = 0.7 could be applied for the entire range of gas Prandtl numbers. When reference temperatures were computed for Pr = 1, they were found to be in excellent agreement with those for Pr = 0.7. So, it is felt that Equation [20] is valid for the entire Prandtl-number range of gases.

Test of Reference-Temperature Procedure. Having established a reference-temperature rule for gas A, we now propose to check whether it can be used for predicting variable-property heat-transfer results for other gases.

We proceed in the following way: For each of the gases of Table 1 (other than gas A), special cases are selected
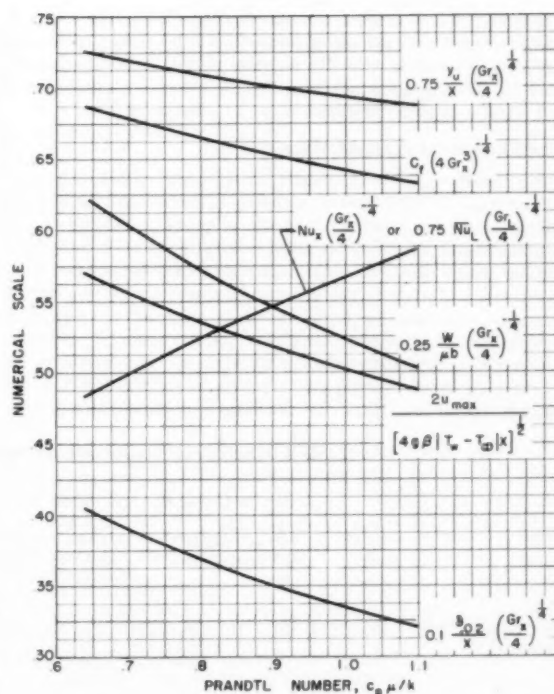
and heat-transfer results computed on the basis of solutions of Equations [8a]. Then for each of these special situations, an alternate heat-transfer result is computed using the reference-temperature rule of Equation [20] in conjunction with the constant-property results, which are plotted in Fig. 3. The comparison between the two sets of heat-transfer results, given in Table 3, provides a test of the reference-temperature procedure. Comparison of the last two columns indicates that the reference-temperature procedure works very well indeed.

Two observations are worthy of mention in connection with the computations of Table 3. The first has to do with the application of the reference-temperature procedure to gas E, which has a temperature-dependent Prandtl number.[3] For this gas, not only were k and ν (of the Nusselt and Grashof numbers) evaluated at

---
[3] Over the temperature range investigated, the Prandtl-number variation was 4.5 per cent.

Table 3 Tests of reference-temperature procedure

Gas	Temp. condition	Pr	Solution of diff. eq.	Reference temp. procedure
		$\mathrm{Nu}_{x,w}/\mathrm{Gr}_{x,w}{}^{1/4}$ or $\overline{\mathrm{Nu}}_{L,w}\left/\dfrac{4}{3}\,\mathrm{Gr}_{L,w}{}^{1/4}\right.$		
B	$T_w/T_\infty = 3$	0.7	0.373	0.370
C	Arbitrary	0.7	0.353	0.353
	Arbitrary	1.0	0.401	0.401
D	$\begin{cases} T_w = 1800 \text{ F abs} \\ T_\infty = 600 \text{ F abs} \end{cases}$	0.7	0.371	0.370
	$\begin{cases} T_w = 600 \text{ F abs} \\ T_\infty = 1800 \text{ F abs} \end{cases}$	0.7	0.335	0.335
E	$\begin{cases} T_w = 1800 \text{ F abs} \\ T_\infty = 600 \text{ F abs} \end{cases}$	$\mathrm{Pr}_w = 0.7$	0.358	0.361
	$\begin{cases} T_w = 600 \text{ F abs} \\ T_\infty = 1800 \text{ F abs} \end{cases}$	$\mathrm{Pr}_w = 0.7$	0.346	0.343

the reference temperature, but also the Prandtl number. Also, it may be noted that the use of Sutherland-type variations for the conductivity and/or viscosity requires that specific values of T_w and T_∞ be assigned before the differential Equations [8a] can be solved; i.e., specific problems must be considered.

Fluid Properties of Liquid Mercury

As already noted in the Introduction, it was initially hoped to study the variable-property problem in several liquid metals. However, it was observed that the property variations of the various liquid metals of technical interest were sufficiently dissimilar that separate investigations appeared to be necessary for each one. Moreover, the computing time required to obtain numerical solutions for liquid metals is an order of magnitude greater than that required for gases. So, it was found necessary to limit the current study to one liquid metal; namely, mercury.

Property data for liquid mercury were taken from the "Liquid-Metals Handbook" (6). To facilitate numerical integrations,

A plot of the constant-property heat-transfer results is given in Fig. 4(a) for the Prandtl-number range appropriate to liquid mercury. After making a few trials, it is found that by evaluating k, ν, β, and Pr at

$$T_r = T_w - 0.3(T_w - T_\infty) \dots \dots \dots \dots [23]$$

the constant-property heat-transfer results coincide with those of Equation [22]. It is emphasized that this same reference temperature is found for both variable-property cases studied here.

Next, it may be inquired as to how well the film temperature T_f might serve as an alternate reference temperature. The heat transfer calculated for cases I and II by evaluating the constant-property results of Fig. 4(a) at the film temperature T_f agree quite well with Equation [22], certainly within the range of most engineering requirements. The same statement applies when all the properties (including β) are evaluated at $T_r = T_w - 0.38$ $(T_w - T_\infty)$, which was the reference temperature derived for gases.

Table 4 Coefficients for polynomial representations of properties of liquid mercury

	k, Btu/hr ft deg F	μ, lb/ft hr	c_p, Btu/lb deg F	ρ, pcf
a_0	4.47924	4.34620	$0.334620(10^{-1})$	851.514
a_1	$0.830958(10^{-2})$	$-0.991162(10^{-2})$	$-0.393353(10^{-5})$	$-0.864880(10^{-1})$
a_2	$-0.380163(10^{-5})$	$0.179060(10^{-4})$	$0.344649(10^{-8})$	$0.986194(10^{-3})$
a_3	0	$-0.127524(10^{-7})$	0	$-0.592566(10^{-5})$

polynomial representations of the following form were fitted to the data

$$\sum_{n=0}^{3} a_n t^n \dots \dots \dots \dots \dots \dots [21]$$

where t is in degrees Fahrenheit. The coefficients a_n are given in Table 4.

Heat-Transfer Results for Liquid Mercury

First, based on numerical solutions of Equations [8], heat-transfer results are given for special cases. Then these findings are utilized in determining reference temperatures.

Results for Specific Cases. For liquid mercury, the buoyancy force appearing in the first of Equations [8] does not simplify as it did for gases. Hence it is necessary to deal with Equations [8] as they stand. Further, the nature of the property variations of liquid mercury requires that t_w and t_∞ be specified separately; i.e., that specific problems be considered. The two cases chosen for study will be described. In both instances $t_w > t_\infty$.

$$\left. \begin{array}{l} t_w = 600 \text{ F} \\ t_\infty = 100 \text{ F} \\ \text{Pr}_w = 0.0083 \\ \text{Pr}_\infty = 0.022 \end{array} \right\} \text{case I} \qquad \left. \begin{array}{l} t_w = 450 \text{ F} \\ t_\infty = 150 \text{ F} \\ \text{Pr}_w = 0.010 \\ \text{Pr}_\infty = 0.019 \end{array} \right\} \text{case II}$$

On the basis of the rather lengthy numerical solutions of the differential Equations [8], the heat-transfer results are found to be

$$\text{Nu}_{x,w}/\text{Gr}_{x,w}^{1/4} = \overline{\text{Nu}}_{L,w} \bigg/ \sqrt[4]{\frac{4}{3}} \text{Gr}_{L,w}^{1/4} = \begin{cases} 0.0501, \text{ case I} \\ 0.0556, \text{ case II} \end{cases} \dots [22]$$

where the Grashof number is given by Equation [13].

Reference Temperature Relation. Now, it may be inquired as to whether there is some reference temperature rule which will cause the constant-property results to coincide with those of Equation [22].
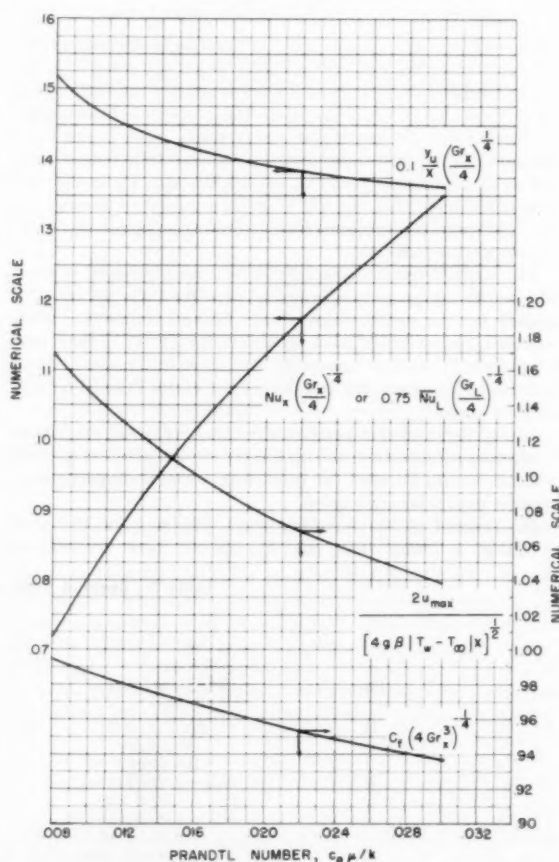

Fig. 4(a) Results of constant-property analysis for Prandtl-number range of liquid mercury
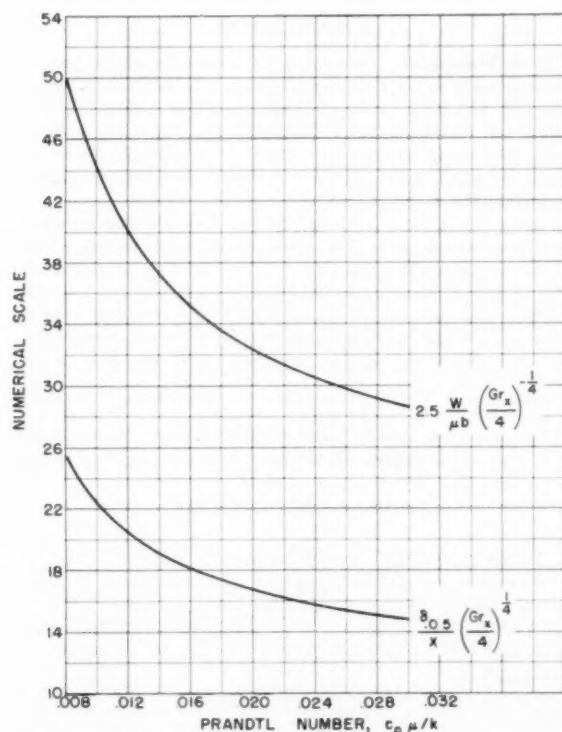
Fig. 4(b) Results of constant-property analysis for Prandtl-number range of liquid mercury

Experimental Verification

A complete survey of available experimental heat-transfer data for free convection on a vertical plate is presented in reference (1). It is noted there that most of the experiments were carried out in air.

For air, the conditions of most tests were such that $1 < T_w/T_\infty \leq 1.5$. These experimenters took no note of variable-property effects. This action is justified by the results of the present analysis; for example, see Fig. 2. In one experiment, that of Weise (7), a value of T_w/T_∞ of 2.2 was achieved. Unfortunately, Weise's apparatus, constructed for horizontal plate tests, was not well suited for vertical-plate studies.

For liquid mercury, Saunders' (8) work represents the only experiment on a vertical plate. His tests were carried out at very small temperature differences, and hence variable-property effects did not enter.

So, it appears that there are currently no data available to check the findings of the analysis presented here.

Results for Boundary-Layer Thickness and Velocity Parameters

While the heat transfer is by far the result of greatest practical importance, there are other quantities which may be of interest. Several of these will be described briefly. Following the descriptive paragraphs, results derived from the constant-property analysis are given, and then tables of reference temperatures are supplied for extending these constant-property results to variable property situations.

(a) *Boundary-layer thickness, δ_i.* The thickness of the boundary layer is by no means a precise concept, and its definition is somewhat arbitrary. The distance from the plate surface at which

$T - T_\infty$ has shrunk to a small fraction i of the over-all temperature difference, $T_w - T_\infty$, is used here to define a boundary-layer thickness. This definition, while satisfactory for Prandtl numbers below unity, should not be used for high Prandtl numbers.

(b) *Maximum velocity, u_{max}.* The vertical velocity u (parallel to the plate surface) takes on zero values both at the surface and at $y = \infty$ and hence achieves a maximum between.

(c) *Location of maximum velocity, y_u.* The distance from the plate surface at which the velocity maximum occurs is denoted by y_u.

(d) *Friction coefficient, c_f.* The coefficient of friction provides a dimensionless presentation of the wall shear stress τ_w in the following manner

$$c_f \equiv \frac{\tau_w}{\rho \left(\dfrac{\nu}{x}\right)^2} = \frac{\left[\mu \dfrac{\partial u}{\partial y}\right]_{y=0}}{\rho \left(\dfrac{\nu}{x}\right)^2} \qquad \ldots\ldots\ldots\ldots [24]$$

The quotient (ν/x) plays the role of a characteristic velocity.

(e) *Flow rate, W.* The upward flow generated by the free-convection forces is given by

$$W = b \int_0^\infty \rho u \, dy \ldots\ldots\ldots\ldots\ldots [25]$$

where b is the plate width.

Results From Constant-Property Analysis. Utilizing the tabulated constant-property solutions of Ostrach (2) in conjunction with those of reference (1), results for the quantities described in paragraphs (a) through (e) have been computed. A presentation of these findings for the Prandtl-number ranges appropriate to gases and to liquid mercury is made, respectively, in Figs. 3 and 4. Heat-transfer results also are shown. The Grashof number is given by Equation [13a].

Originally, it was decided to define the boundary-layer thickness as the distance from the plate where $(T - T_\infty)/(T_w - T_\infty) = 0.02$. For the Prandtl-number range of gases, there was no difficulty evaluating such a definition from the available numerical solutions. But, for the low Prandtl-number range, the numerical solutions were less precise at large distances from the wall, and it was necessary to use a thickness based on $(T - T_\infty)/(T_w - T_\infty) = 0.05$. It is estimated that for the Prandtl-number range of mercury, $\delta_{0.05} = 1.25 \delta_{0.04}$.

Variable-Property Results; Reference Temperatures. In a manner identical to that outlined for the heat transfer, the findings of the constant property analysis for the quantities described above may be extended to the variable property situation. It only remains to present appropriate reference temperatures which are needed to evaluate the results of Figs. 3 and 4.

Such reference temperatures, computed using the variable-property solutions which have been obtained here for gases and for liquid mercury, are listed in Tables 5(a) and 5(b). The heat-transfer reference temperatures, already given by Equations [20] and [23], are included for completeness.

Table 5(a) Reference-temperature relations for gases. (For use with Fig. 3.)

	$(T_r - T_w)/(T_\infty - T_w)$
h	0.38
$\delta_{0.02}$	0.67
u_{max}	See below
y_u	0.24
c_f	0.10
W	0.85

$\beta = \dfrac{1}{T_\infty}$

To find u_{max} for gases, evaluate the Prandtl number in Fig. 3

at the mean of the extreme values of Pr in the particular problem under consideration.

Table 5(b) Reference-temperature relations for liquid mercury. (For use with Figs. 4a and 4b.)

	$(T_r - T_w)/(T_\infty - T_w)$
h	0.30
$\delta_{0.05}$	0.60
u_{max}	0
y_u	0
c_f	0.10
W	0.40

Conclusion

From the findings reported here, it appears that free-convection heat-transfer under variable-property conditions can be computed quickly and accurately by using the constant-property results in conjunction with reference-temperature relations. For gases and liquid mercury, the reference-temperature relations are given by Equations [20] and [23], respectively. Further, it may be observed that the film temperature appears to serve as an adequate reference temperature (with $\beta = 1/T_\infty$ for gases) for most engineering purposes.

Reference-temperature relations for use in computing boundary-layer thickness and velocity parameters for variable-property conditions are given in Tables 5(a) and 5(b).

The results obtained here are based on laminar boundary-layer theory and are expected to be applicable within the range of validity of this model. The extension of the results to situations outside this range is uncertain.

Acknowledgment

It is a pleasure to acknowledge the guidance of Prof. Howard W. Emmons of Harvard University.

Bibliography

1 "Free Convection With Variable Properties and Variable Wall Temperature," by E. M. Sparrow, PhD thesis, Harvard University, Cambridge, Mass., May, 1956.

2 "An Analysis of Laminar Free-Convection Flow and Heat Transfer About a Flat Plate Parallel to the Direction of the Generating Body Force," by S. Ostrach, NACA Report 1111, 1953 (supersedes NACA TN 2635).

3 "Heat Interchange Under Conditions of Free Laminar Motion of a Gas With Variable Viscosity at a Vertical Wall," by A. A. Tanaev, *Zhurnal Tekhnicheskoi Fiziki*, vol. 26, December, 1956, p. 2714.

4 "Thermal Conductivity and Prandtl Number of Air at High Temperatures," by I. Glassman and C. F. Bonilla, *Chemical Engineering Progress Symposium Series*, vol. 49, no. 5, 1953, p. 153.

5 "Tables of Thermal Properties of Gases," by Joseph Hilsenrath, et al., NBS Circular 564, November 1, 1955.

6 "Liquid-Metals Handbook," edited by R. N. Lyon, Atomic Energy Commission, second edition, June, 1952.

7 "Warmeubergang durch freie Convection an quatratischen Platten," by R. Weise, *Forschung auf dem Gebiete des Ingenieurwesens*, Bd. 6, 1935, p. 289.

8 "Natural Convection in Liquids," by O. A. Saunders, *Proceedings of the Royal Society of London*, series A, vol. 172, no. 948, July 19, 1939, pp. 55–71.

Discussion

J. R. Moszynski.[9] The authors are to be complimented on their attempt to tackle a very laborious, albeit important, problem and their results are particularly gratifying inasmuch as they show that the comparatively simple constant-property analysis may, with small modifications, yield accurate heat transfer predictions under conditions where the variations of thermodynamic properties are of importance.

While it is perhaps surprising that an exact solution for gas A can be approximated so well by the use of a single reference temperature over a wide range of T_w/T_∞, it may be expected that in the particular cases presented in Table 3 this same reference temperature used in conjunction with the constant-property solutions will give fairly accurate results. For the case of $T_w/T_\infty = 3$ a comparatively simple computation will show that the expression $\overline{\mathrm{Nu}}_{L,w} / \frac{4}{3} \mathrm{Gr}_{L,w}{}^{1/4}$ is rather insensitive to the exponent n of the power law used to represent the variation of thermal conductivity and viscosity. Thus for example if $k \sim T^{1/4}$; $\mu \sim T^{1/2}$

$$\overline{\mathrm{Nu}}_{L,w} / \frac{4}{3} \mathrm{G}_{L,w}{}^{1/4} = 0.379$$

and from Table 2 we obtain 0.368 for $n = 3/4$.

Furthermore it may be shown that for the cases of $n = 1/2$ and $n = 2/3$ the application of the usual film temperature $T_f = T_w - 0.5(T_w - T_\infty)$ with $\beta = 1/T_\infty$ yields values not too widely divergent from those obtained with the authors' reference temperature rule.

For the region of high absolute temperatures, investigated in the paper, the Sutherland-type equations of gases D and E differ but little from Maxwell's equations for viscosity and thermal conductivity and, in the light of the foregoing, the good agreement shown in Table 3 is to be expected.

It should be worth while, however, to investigate the accuracy of the reference temperature rule for similarly high values of T_w/T_∞, but at lower temperatures where the Sutherland equations diverge more from the simple approximation $\mu \sim T^n$; $k \sim T^n$.

The writer would also like to know whether the reference temperature relations of Table 5a have been tested for several types of property variation similarly to the heat transfer results.

Authors' Closure

The authors wish to extend their thanks to Dr. Moszynski for his interest and comments. The further tests of the reference temperature procedure for the very low temperature range, as suggested by him, would be interesting; but such a study is not feasible at present. With regard to the question about the reference temperatures of Table 5(a), all were derived and tested in a manner similar to that described for the heat transfer. Further details may be found in chapter 4 of reference (1).

9 Division of Engineering, Brown University, Providence, R. I.

Laminar Film Condensation of Pure Saturated Vapors on Inclined Circular Cylinders

By KAMAL-ELDIN HASSAN[1] AND MAX JAKOB[2]

Laminar film condensation of pure saturated vapors on inclined circular cylinders is treated analytically. The analysis is based on, and made under the same assumptions of, Nusselt's theory of film condensation. Experiments were performed to check the analytic results.

In the analysis, the problem is expressed in terms of dimensionless groups and the solution gives, in these terms, the local film thickness at any point on the surface of a cylinder. From the results of the analysis one can obtain the local film coefficient, and hence a mean film coefficient for the entire surface of any cylinder at any inclination.

The analysis shows that every tube has an optimum inclination at which the rate of condensation on the entire tube is a maximum.

The experiments were carried out by condensing steam on the inside surface of a thick copper tube $1\frac{1}{4}$ in. diam and about $8\frac{1}{8}$ in. long. The actual film coefficients obtained during the experiments were 28 to 100 per cent higher than those predicted by the theoretical analysis. The deviation from the theory increased with the inclination of the tube and the temperature difference between the tube surface and the steam.

Nomenclature

The following nomenclature is used in the paper:

A = area, sq ft

$A_{i,j}, \ldots$ = factors in Equation [28]

C = const

$\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ = unit vectors in the x_1, x_2, and x_3-directions

H = width of mesh in the X-direction, Fig. 14

h = local coefficient of heat transfer, Btu/hr sq ft deg F

h_M = mean coefficient of heat transfer on entire surface of cylinder, Btu/hr sq ft deg F

h_m = mean coefficient of heat transfer on a ring of infinitesimal width on cylinder surface, Btu/hr sq ft deg F

h_{Ma} = actual mean coefficient of heat transfer from entire surface of a tube, Btu/hr sq ft deg F

K = width of mesh in the ϕ-direction, Fig. 14

k = conductivity of condensate, Btu/hr ft deg F

L = length of cylinder, ft

L_{st} = starting length, ft

$\dot{m}'$ = rate of condensation per unit length of an infinitely long cylinder, lb/hr ft

$\mathbf{N}_c$ = unit vector normal to surface of cylinder

$\mathbf{N}_h$ = unit vector normal to horizontal plane

$\mathbf{N}_{int}$ = unit vector normal to plane of $\mathbf{N}_c$ and $\mathbf{N}_h$

N_{Re} = Reynolds number, $4v_m Y\rho/\mu$

q = rate of heat flow, Btu/hr

r = cylinder radius, ft

s = distance along horizontal in x_1, x_2-plane, a parameter in Equation [2]

t = temperature, deg F

t_s = temperature of condensing surface, deg F

t_v = temperature of saturated vapor, deg F

u = distance along cylinder axis, a parameter in Equation [1]

v = distance along x_2-axis, a parameter in Equation [2]

v_m = mean velocity at any cross section of condensate film, ft/hr

X = reduced distance, $x_1/r \tan \alpha$

X_L = $L/r \tan \alpha$

x_1, x_2, x_3 = co-ordinate axes, shown in Fig. 2

Y = thickness of condensate film, ft

Z = $\gamma^2 \lambda \cos \alpha Y^4/3\mu k\, r\, \Delta t$

α = angle of inclination of cylinder with horizontal

γ = specific weight, lb$_F$/cu ft

Δt = temperature difference, $t_v - t_s$, deg F

$(\Delta t)_m$ = mean temperature difference for a nonisothermal surface, deg F

$$= \frac{1}{A} \int\!\!\int \Delta t\, dA$$

θ = angle between a plane tangent to cylinder and horizontal plane

λ = latent heat of condensation, Btu/lb$_F$

μ = dynamic viscosity, lb$_F$ hr/sq ft

ρ = density, lb$_M$/cu ft

ϕ = angular position from top element of cylinder

ψ = angle between a momentary droplet path and a cross section of cylinder

Analysis

Assumptions

Nusselt's (1)[3] basic idea of filmwise condensation[4] is that the heat energy liberated by the condensing vapor is conducted through the film to the cooler surface. His basic assumptions which are retained in the present analysis, seem never to have been stated clearly. Hence it might be of some benefit to state the assumptions and conditions pertinent to the present problem. These assumptions are:

1 The vapor is pure, dry, and saturated.

2 The condensate-film flow is laminar.

3 The vapor at the vapor-liquid interface is stagnant and the shear stress is negligible at this interface.

4 The wall temperature is uniform.

5 The liquid-solid and liquid-vapor interfaces are smooth.

[1] Lecturer, Mechanical Engineering Department, Faculty of Engineering, University of Khartoum, Khartoum, Sudan. Assoc. Mem. ASME.

[2] Late Research Professor of Mechanical Engineering, Illinois Institute of Technology, Chicago, Ill. Mem. ASME.

[3] Numbers in parentheses refer to the Bibliography at the end of the paper.

[4] In discussing Nusselt's work, reference will be made to Jakob (2) who presented most of it.

6 The velocity distribution at any point on the cylinder surface is the same as that in a fully developed isothermal film flowing on a plane tangent to the surface at that point. Under such conditions the effect of acceleration in the film is neglected. Also, the effect of the curvature of the surface on the velocity distribution in the film is neglected.

7 The curvature of the surface is large enough that the effect of capillary forces may be neglected.

8 The liquid temperature at the liquid-vapor interface is that of the saturated vapor.

9 The convective heat transfer along the condensate film is neglected. Thus considering the part of the condensate film in Fig. 1, it is assumed that $q_3 = q_4$. Consequently, $q_1 = q_2$. Further, from the previous assumption, q_1 is equal to the latent heat liberated by the condensing vapor.

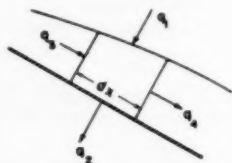10 The physical properties of the condensate are constant.

Fig. 1   Element in condensate film

Droplet Path

In accordance with the foregoing assumptions, a droplet of condensate at any point on the surface is acted upon by the force of gravity which is in equilibrium with the viscous forces. Since the motion is constrained by the surface, a droplet at any point moves, momentarily, along the line of intersection of the plane tangent to the cylinder at this point and the vertical plane normal to this tangent plane.

Consider the case of a circular cylinder of radius r inclined at an angle α to the horizontal. With the co-ordinate axes as shown in Fig. 2, the parametric equations of the cylinder and the horizontal plane are, respectively

$$\left. \begin{array}{l} x_1 = u \\ x_2 = r \sin \phi \\ x_3 = r \cos \phi \end{array} \right\} \quad \ldots \ldots \ldots \ldots \ldots \ldots [1]$$

and

$$\left. \begin{array}{l} x_1 = s \cos \alpha \\ x_2 = v \\ x_3 = s \sin \alpha \end{array} \right\} \quad \ldots \ldots \ldots \ldots \ldots \ldots [2]$$

Using $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ as unit vectors in the x_1, x_2, and x_3-directions, respectively, the unit normal vectors to the cylinder surface and the horizontal plane would be given, respectively, by

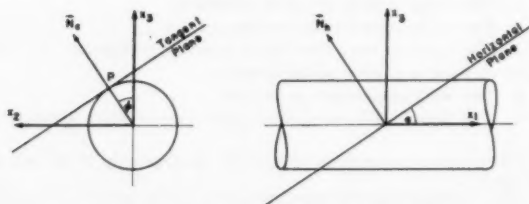$$\mathbf{N}_c = \sin \phi \, \mathbf{e}_2 + \cos \phi \, \mathbf{e}_3 \ldots \ldots \ldots \ldots \ldots [3]$$

Fig. 2   Co-ordinate system with denotations

$$\mathbf{N}_h = -\sin \alpha \, \mathbf{e}_1 + \cos \alpha \, \mathbf{e}_3 \ldots \ldots \ldots \ldots \ldots [4]$$

The angle θ between the momentary path of a droplet at a point P on the cylinder and the horizontal is the angle between the plane tangent to the cylinder at the point P and the horizontal plane. The angle θ is, therefore, equal to the angle between the two normal unit vectors $\mathbf{N}_c$ and $\mathbf{N}_h$ and, hence, is obtained from

$$\cos \theta = \cos \alpha \cos \phi \ldots \ldots \ldots \ldots \ldots [5]$$

Now if ψ denotes the angle between the momentary droplet path at P and a cross section of the cylinder, it will be identically the angle between the $x_2 x_3$-plane and the line of intersection of the tangent plane at P and the plane of $\mathbf{N}_c$ and $\mathbf{N}_h$. Since the $x_2 x_3$-plane and the plane of $\mathbf{N}_c$ and $\mathbf{N}_h$ are both normal to the tangent plane, therefore ψ is equal to the angle between the x_1-axis and the unit normal $\mathbf{N}_{int}$ to the plane of $\mathbf{N}_c$ and $\mathbf{N}_h$. The unit normal $\mathbf{N}_{int}$ is parallel to the line of intersection of the tangent and the horizontal planes and is given by

$$\mathbf{N}_{int} = \frac{\cos \alpha \sin \phi \, \mathbf{e}_1 - \sin \alpha \cos \phi \, \mathbf{e}_2 + \sin \alpha \sin \phi \, \mathbf{e}_3}{(\sin^2 \alpha + \cos^2 \alpha \sin^2 \phi)^{1/2}} \ldots [6]$$

Therefore the angle ψ is obtained from

$$\cos \psi = \sin \phi / (\tan^2 \alpha + \sin^2 \phi)^{1/2} \ldots \ldots \ldots [7]$$

This relation is now employed to develop the parametric equations of the droplet path. Referring to Fig. 3

$$dx_1 = r \, d\phi \tan \psi$$

$$= r \frac{\tan \alpha}{\sin \phi} d\phi$$
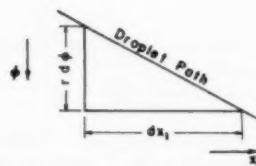
Fig. 3   Momentary droplet path

Integrating

$$x_1 = r \tan \alpha \ln \tan \frac{\phi}{2} + C \ldots \ldots \ldots \ldots [8]$$

where C is the constant of integration.

Therefore the family of space curves along which droplets on the surface of an inclined cylinder move is given by the parametric equations

$$\left. \begin{array}{l} x_1 = r \tan \alpha \ln \tan \frac{\phi}{2} + C \\ \\ x_2 = r \sin \phi \\ \\ x_3 = r \cos \phi \end{array} \right\} \quad \ldots \ldots \ldots \ldots [9]$$

where C is the parameter of the family.

Fig. 4 shows a droplet path on half of the developed surface of the cylinder. The curve shown passes through the point $(0, r, 0)$; it is the curve for $C = 0$.

It should be noted that any droplet path is asymptotic to the lines $\phi = 0$ and π.

The Differential Equation

Consider a rectangular element on the surface of the cylinder as
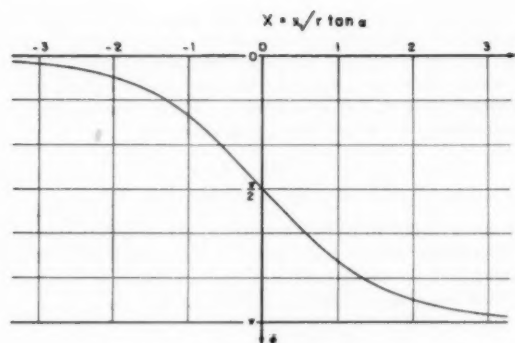
Fig. 4   Droplet path through $(0, r, 0)$
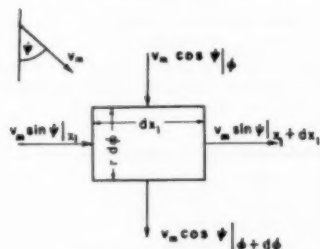
Fig. 5   Surface element

shown in Fig. 5.   The condensate with thickness Y enters this element with a mean velocity v_m making an angle ψ with the lines of constant z_1.   From the assumptions 5, 8, 9, and 10, the rate of heat transferred by conduction through the film is equal to the rate at which heat is liberated by the vapor condensing on the element.   Therefore

$$\frac{k\,r}{Y}\,\Delta t\,dz_1 d\phi = \gamma\lambda\left[\frac{\partial}{\partial\phi}\,(v_m\,Y\cos\psi\,dz_1)d\phi\right.$$
$$\left. + \frac{\partial}{\partial x_1}\,(v_m\,Y\,r\sin\psi\,d\phi)dz_1\right]\ldots\,[10]$$

Nusselt's expression for the mean velocity of an isothermal liquid film on an inclined flat plate, with no shear stress at the liquid-gas interface is

$$v_m = \frac{\gamma}{3\mu}\,Y^2\sin\theta\ldots\ldots\ldots\ldots\,[11]^{\text{[5]}}$$

From assumption 6, this value of v_m is the same as that in the condensate film.   Therefore substituting the value of v_m from Equation [11] in Equation [10] and using Equations [5] and [7], Equation [10] reduces to

$$\frac{3\mu k r\Delta t}{\gamma^2\lambda} = Y^4\cos\alpha\cos\phi + 3\,Y^3\cos\alpha\sin\phi\frac{\partial Y}{\partial\phi}$$
$$+ 3\,Y^3\,r\sin\alpha\frac{\partial Y}{\partial x_1}\ldots\,[12]$$

Placing
$$Z = \frac{\gamma^2\lambda\cos\alpha}{3\,\mu\,k\,r\,\Delta t}\,Y^4\ldots\ldots\ldots\ldots\,[13]$$

$$X = x_1/r\tan\alpha\ldots\ldots\ldots\ldots\,[14]$$

Equation [12] simplifies to

$$\frac{\partial Z}{\partial X} + \sin\phi\frac{\partial Z}{\partial\phi} = \frac{4}{3}\,(1 - Z\cos\phi)\ldots\ldots\,[15]$$

[5] See, for example, reference (2), equation [30–14], p. 663.

Boundary Conditions

On a semi-infinite cylinder in the range of $0 \leq x_1 \leq \infty$, the condensate starts with zero thickness at the uppermost edge, therefore

$$Z = 0 \quad\text{at}\quad X = 0\ldots\ldots\ldots\ldots\,[16]$$

The second boundary condition expresses the symmetry of the condensate film and its smoothness at the uppermost and lowermost elements of the cylinder, therefore

$$\frac{\partial Z}{\partial\phi} = 0 \quad\text{at } \phi = 0\text{ and }\pi\ldots\ldots\ldots\,[17]$$

Special Cases

An inclined cylinder is the general case for the following special ones:

Inclined Flat Plate.   Substituting $r = \infty$ in Equation [12] reduces it to the equation for condensation on an inclined flat plate.

Vertical Cylinder.   Substituting $\alpha = \pi/2$ in Equation [12] reduces it to that for condensation on a vertical plate[6] which represents also the case of a vertical cylinder.

Horizontal Cylinder.   Nusselt's differential equation representing condensation on a horizontal cylinder[7] can be obtained by substituting $\alpha = 0$ in Equation [12].

An Inclined Cylinder of Infinite Length.   Whereas the film thickness on a semi-infinite plane increases without limit, the condensate on the semi-infinite inclined cylinder approaches a limiting thickness as X increases, except at $\phi = \pi$, as will be seen.   The reason for the finite thickness is that a droplet on the surface of the cylinder can always move by virtue of the gravitational forces to a lower point on the surface.   Since the film is finite at $X = \infty$

$$\frac{\partial Z}{\partial X} = 0 \quad\text{at } X = \infty\ldots\ldots\ldots\ldots\,[18]$$

Hence for large distances from the starting edge of the cylinder, Equation [15] reduces to

$$\frac{3}{4}\sin\phi\frac{\partial Z}{\partial\phi} = 1 - Z\cos\phi\ldots\ldots\ldots\,[19]$$

This, except for the definition of Z, is the differential equation for condensation on a horizontal cylinder.   The solution of this equation was originally tabulated by Nusselt and later (and much more accurately) by Abramowitz (3).

This shows that for an infinitely long inclined cylinder the condensate film has the same geometrical shape as that on a horizontal cylinder.   Further, from the definition of Z, namely Equation [13], the film thickness at any point changes inversely with the fourth root of $\cos\alpha$ if all other quantities remain fixed.   Since, from the basic assumptions, the heat is transferred by conduction through the film, the rate of condensation is inversely proportional to the film thickness.   Therefore the rate of condensation $\dot{m}'(\alpha)$ per unit length on an infinitely long cylinder inclined at the angle α to the horizontal, is proportional to the one fourth power of $\cos\alpha$; that is

$$\dot{m}'(\alpha) = \dot{m}'(0)\cos^{1/4}\alpha\ldots\ldots\ldots\ldots\,[20]$$

This relation is important since, as will be shown later, the condensate film after a short distance from the starting edge of the cylinder (a distance that may be considered as a "starting length") will be practically the same as that at infinity.   Hence under the conditions undertaken in this analysis, the larger part of a

[6] See, for example, reference (2), equation [30–20], p. 665.

[7] See, for example, ibid., equation [30–39], p. 668.

condenser tube of the usual practical proportions behaves as an infinitely long cylinder.

Analytic Results

The differential equation, Equation [15], was solved numerically in the manner described in the Appendix. The results of this solution are condensed in Table 1 and an estimation of their accuracy is contained in the Appendix.

The numerical calculations were carried out to $X = 3.2$ only because, as will be shown later, the conditions at this cross section are, for all practical purposes, very near to those at $X = \infty$.

From the values of Z obtained numerically one can obtain, from Equation [13], the thickness of the condensate film at different points on the cylinder surface. Fig. 6 shows the variation of the film-thickness along the cylinder surface for different values of ϕ. Fig. 7 shows the condensate-film thickness at different cross sections along the cylinder.
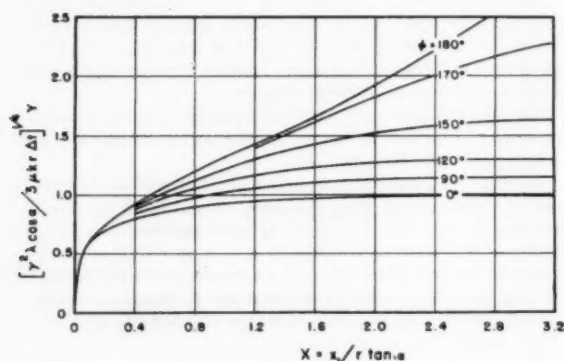

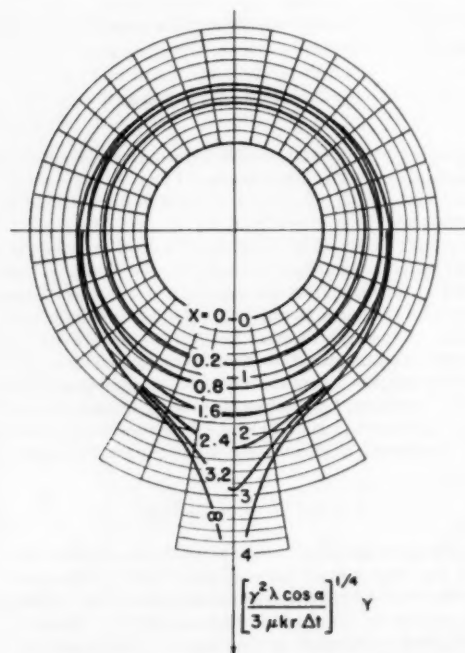Fig. 6 Film thickness at different angular positions

Fig. 7 Polar diagram of dimensionless film thickness at different reduced distances from uppermost end of cylinder

It can be noted in Fig. 6 that for any angular position $\phi \leq 150$ deg, the thickness of the condensate film approaches its limiting value at $X = \infty$ within a rather short reduced distance (X) which lies within the range covered by the numerical calculations.

Further, from the values of Z one can obtain local values of the film coefficient h. From the basic assumptions

$$h = \frac{k}{Y} = \left[\frac{\gamma^2 \lambda \; k^3 \cos \alpha}{3 \; \mu \; r \; \Delta t} \; \frac{1}{Z} \right]^{1/4} \qquad [21]$$

A mean coefficient of heat transfer h_m for an infinitesimal ring on the cylinder surface is a function of the reduced distance X of the ring from the starting end. It can be obtained as the mean of the local values at this distance. Thus for $X = X_1$

$$h_m = \frac{1}{\pi} \int_0^\pi h(X_1, \phi) d\phi \qquad [22]$$

The integral was evaluated numerically by Simpson's two-thirds rule. Values of h_m so obtained are plotted in Fig. 8. They start from infinity at $X = 0$ and decrease to a constant value at $X = \infty$. This constant value is equal to $0.8045 \; (\gamma^2 \lambda k^3 \cos \alpha / 3 \; \mu \; r \; \Delta t)^{1/4}$.
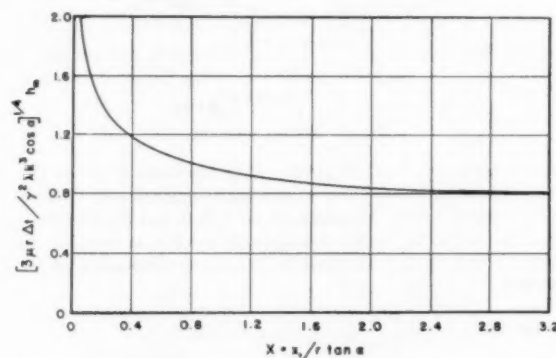
Fig. 8 Mean coefficient of heat transfer at different reduced distances from uppermost end of cylinder

It should be noted that, taking into account the error involved in the numerical calculations, the mean film coefficient at $X = 3.2$ is within less than 2 per cent from its value at $X = \infty$ (4).

From the foregoing considerations, one can consider the part of the tube with $X \leq 3.2$ as a starting length beyond which the condensation is approximately equivalent to that on an infinitely long cylinder. This starting length given by

$$L_{st} = 3.2 \; r \tan \alpha \qquad [23]$$

is quite small in practical cases particularly for small values of α.

The mean film coefficient h_M on the *entire* surface of an inclined cylinder is a function of its reduced length given by

$$X_L = L/r \tan \alpha \qquad [24]$$

This mean film coefficient can be obtained from

$$h_M = \frac{1}{X_L} \int_0^{X_L} h_m(X) dX \qquad [25]$$

Values of h_M were calculated numerically (4) and are plotted in Fig. 9. It is apparent that a cylinder with $X_L = 50$ has a mean film coefficient within less than 2 per cent (taking into account the error in the numerical calculations) of the mean film coefficient for an infinitely long cylinder.
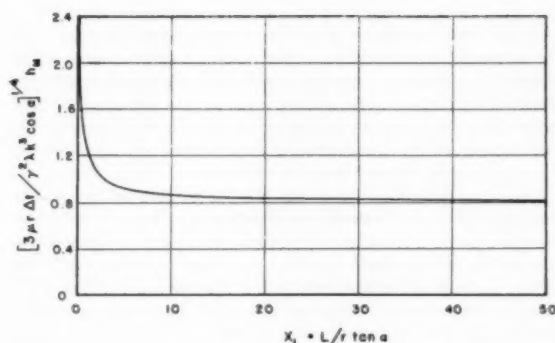
Fig. 9   Mean coefficient of heat transfer on entire surface of inclined cylinders

Optimum Angle of Inclination for Maximum Rate of Condensation

It is noted that the reduced length of a given cylinder is a function of the angle of inclination α. Hence while a given cylinder may be equivalent to an infinitely long cylinder at certain angles of inclination, it may have to be considered quite short at larger angles. The better efficiency of short cylinders (as can be seen from Fig. 9) raises the question whether an optimum inclination exists at which the mean film coefficient h_M is a maximum for cylinders of a fixed ratio L/r. That such an angle exists is shown in Fig. 10.

For an infinitely long cylinder, as pointed out before, h_M is proportional to $\cos^{1/4} \alpha$, therefore its maximum value occurs at $\alpha = 0$. For cylinders of finite length, the angle at which h_M is a maximum increases as the ratio L/r decreases as shown by the dotted line in Fig. 10.
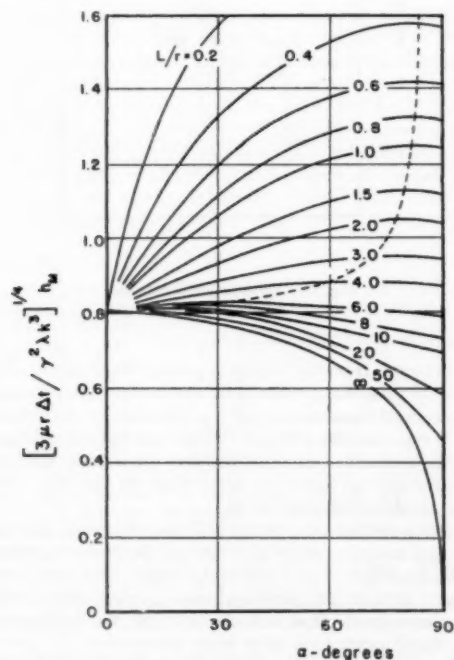

Fig. 10   Variation of mean heat-transfer coefficient on entire surface of geometrically similar cylinders with their inclination

Experiments

Experiments were carried out to check the analysis presented in the previous section rather than to furnish complete experimental data for filmwise condensation on inclined tubes.

In formulating the theoretical analysis, it was not possible, under the basic assumptions, to differentiate between the two cases of condensation on the outside and inside surfaces of a tube. While the results of the analysis are physically possible for condensation on the outside surface of a tube, they are not so for condensation on the inside surface in the vicinity of $\phi = \pi$ for large values of X. In this region, the condensate film cannot rise above its level in the neighboring regions. However, as can be seen from Fig. 7, this discrepancy is limited to a very small region which is, anyway, the least efficient. In view of this fact, the experiments were carried out by condensing pure steam on the inside surface of a thick copper tube. The apparatus in this case is simple and the sources of experimental errors are greatly reduced.

The Apparatus

Fig. 11 shows the general arrangement of the apparatus. Steam is raised in the boiler A and supplied, through lagged connections, to the experimental tube D. The moisture in the steam, if any, is isolated in the separator B and collected in a vessel C. The water level in C is maintained about 6 in. above the end of the hose to act as a safety valve for the apparatus. The cooling water in the annular space E condenses most of the steam that enters the experimental tube D.[8] The condensate is collected in the tube F.
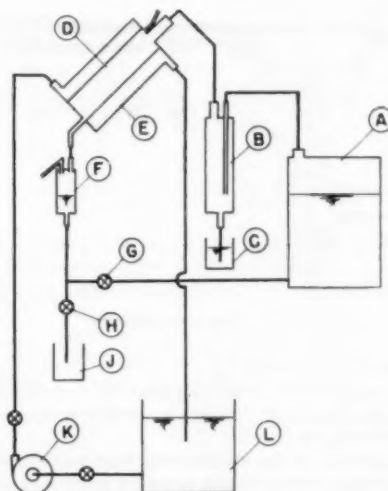
Fig. 11   General arrangement of apparatus

When an experiment is started, the valve H is kept closed and the valve G open. Under such conditions the condensate collected in F flows back to the boiler.

When steady conditions are reached, the valve G is closed and the valve H is opened. The rate of condensation can then be determined by measuring the condensate collected in the graduated cylinder J in a known time interval.

The cooling water is circulated by the pump K. Electric heaters in the reservoir L regulate the temperature of the cooling water.

[8] The remainder of the steam is vented from tops of the experimental tube D and the condensate collecting tube F.

The experimental tube and its water jacket are shown in Fig. 12. The experimental tube M is made of copper. It is $1\frac{1}{4}$ in. ID diameter, $8\frac{1}{2}$ in. long, and $\frac{1}{4}$ in. thick. The ends of the tube are covered with the brass covers N and P. The covers are kept tight by the nuts Q. Rubber gaskets R and V prevent leakage from the annular space to either the experimental tube or the outside. The outer shell of the annular space is made from two 2-in. tees S and T and a nipple U. The cooling water enters through the tube assembly W and leaves from the tube assembly X; both are fitted with thermocouples as shown. The whole assembly is mounted on a panel Y and can rotate about the exit pipe assembly in the bushing Z. Holes for settings of $\alpha = 0, 10, 20, 30, 45, 60, 75,$ and 90 deg are provided. The assembly can be held securely in any desired position by screwing the bolt h in the rod g (which is welded to T), and the nut j against the bushing Z.
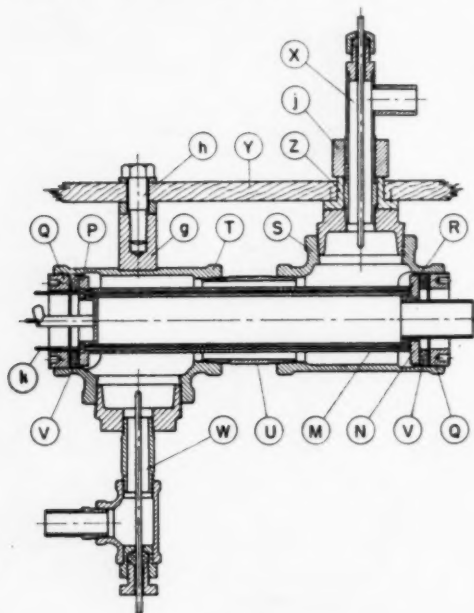

Fig. 12   Experimental tube

Temperature Measurements

All the temperatures were measured by copper-constantan thermocouples connected, through a selector switch, to an accurate potentiometer.

The temperature of the experimental tube wall was measured by four traveling thermocouples in milled grooves spaced at 90 deg around the tube and denoted by k in Fig. 12. The temperatures at 1-in. intervals along the tube were measured by each of these thermocouples. The mean temperature of the tube wall was taken as that corresponding to the arithmetic mean of the 36 readings indicated by these thermocouples.

The steam temperature was measured in the separator. The temperature of the condensate was measured as it leaves the experimental tube. Further, the inlet and outlet cooling-water temperatures were measured also.

Difficulties Encountered

Some difficulties were encountered during the experimentation. Two of these difficulties are discussed here. The first difficulty was in maintaining filmwise condensation. It was impossible with the small rates of condensation used to maintain filmwise con-

densation with pure water. Pure steam condensed in the form of droplets. This was detected by the relatively large fluctuations in the tube-wall temperatures. However, the addition of a small amount of a wetting agent[9] reduced these fluctuations greatly. These fluctuations were always present but their amplitude varied at different points and, after the use of the wetting agent, were mostly, as registered on the galvanometer, within ± 0.06 deg F and never more than double this amount.

The other difficulty was due to the contamination of the condensing surface. Filmwise condensation was difficult to maintain on a badly contaminated surface. Further, the contamination on the surface seems to offer a considerable resistance to the flow of heat, which induces an error in the reading of the surface temperature. For a given temperature-difference reading, a newly cleaned surface condensed about 10 per cent more than the results presented here which were obtained after condensation took place on a cleaned surface for about 24 hr. These results represent more stable conditions and were staggered to check the effect of contamination.

Results

The results are shown in Fig. 13. The abscissa is $(\Delta t)_m$, the difference between the steam temperature and the mean tube temperature. The ordinate is the ratio between the actual mean film coefficient h_{Ma} and the theoretical film coefficient h_M. These film coefficients were based on the mean temperature difference $(\Delta t)_m$.
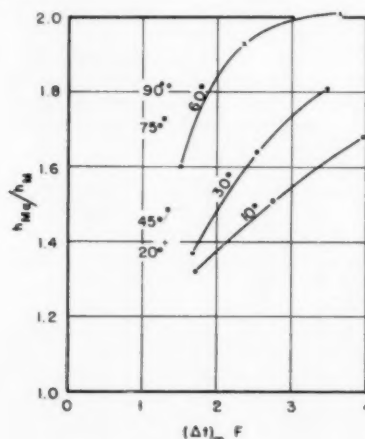
Fig. 13   Experimental results

Discussion

As can be seen from Fig. 13, the apparatus invariably condensed more steam than that predicted by the theory. This increase in the rate of condensation is real as was shown by an analysis of the probable experimental errors (4). This analysis showed that the error due to condensation on surfaces other than the experimental surface can never be more than 10 per cent. Other sources introduce negligible errors.

Another possible cause of the difference between the experimental and analytic results is the deviation, in the experiments, from the condition of an isothermal surface. However, from an analysis of laminar film condensation on nonisothermal surfaces, it seems improbable that such deviation, in the present experiments, would account for these large differences.

It also might be mentioned that the results obtained with the

[9] The wetting agent used was Kodak Photo-Flo.

tube in the horizontal position agree with those cited by Reddie (5). For the vertical position, the present results agree with those of Fragen, as cited by McAdams (6) who tabulated the results of various workers. The experiments of Fragen seem to be the only ones that had been carried out under conditions similar to those of the present tests. For other positions, the only experiments that appear to have been carried out on inclined tubes are those of Tepe and Mueller (7). However, the only inclination they experimented on was 15 deg to the horizontal. Their results cannot be compared with the present ones since they used large temperature differences and appreciable steam velocities in the experimental tube.

The present results show the following tendencies:

1 For a given inclination, the ratio h_{Ma}/h_M increases with $(\Delta t)_m$.

2 For a given temperature difference, the ratio h_{Ma}/h_M increases with the tube inclination α.

3 For a given inclination, the quantity $h_{Ma}(\Delta t)_m^{1/4}$ increases with $(\Delta t)_m$. Theoretically, this quantity should remain constant.

The deviation of the measured film coefficient from the theoretical value is believed to be due to ripples at the condensate-film surface. These ripples are believed to be the cause of the small fluctuations registered by the tube-wall thermocouples. The rippling of flowing liquid films was detected by various investigators. Kirkbride (8) found the ripples in films with Reynolds numbers, $N_{Re} = 4\,v_m\,Y\rho/\mu$, as low as 6.5. The presence of ripples, naturally, increases the mean film coefficient. For example, calculations based on instantaneous film thicknesses recorded by Dukler and Bergelin (9), for $N_{Re} = 785$ showed that the film coefficient of the rippling film is larger than that of an equivalent smooth film by nearly 50 per cent.

Conclusions

The deviation of the experimental results from those obtained analytically is quite large; however, in agreement with other experimental results obtained under similar conditions. This deviation is believed to result from the rippling of the condensate film. It appears, therefore, that a better understanding of the effect of the ripples on the local film coefficient is necessary for a better prediction of the rate of laminar-film condensation on any surface.

From the purely engineering viewpoint, the inclined tube provides a promising condensing surface which seems not to have been investigated thoroughly as yet. In this respect, more comprehensive experimental data are necessary.

Bibliography

1 "Die Oberflachenkondensation des Wasserdampfes," by W. Nusselt, Zeitschrift des Vereines deutscher Ingenieure, vol. 60, 1916, pp. 541–546 and 569–575.

2 "Heat Transfer" by M. Jakob, John Wiley & Sons, Inc., New York, N. Y., 1949, vol. 1, pp. 658–680.

3 "Tables of the Functions $\int_0^\phi \sin^{1/3} x\, dx$ and $\frac{4}{3} \sin^{-4/3} \phi \int_0^\phi \sin^{1/3} x\, dx$," by M. Abramowitz, Journal of Research of the National Bureau of Standards, vol. 47, 1951, pp. 288–290.

4 "Laminar Film Condensation of Pure Saturated Vapors on Inclined Circular Cylinders," by K. Hassan, unpublished PhD Thesis, Illinois Institute of Technology, Chicago, Ill., 1955.

5 "Heat Transfer Coefficients for Vapors Condensing on Horizontal Tubes," by W. A. Reddie, unpublished PhD thesis, Illinois Institute of Technology, Chicago, Ill., 1948.

6 "Heat Transmission," by W. H. McAdams, McGraw-Hill Book Co., Inc., New York, N. Y., 3rd edition, 1954, p. 333.

7 "Condensation and Subcooling Inside an Inclined Tube," by J. B. Tepe and A. C. Mueller, Chemical Engineering Progress, vol. 43, 1947, pp. 267–278.

8 "Heat Transfer by Condensing Vapors on Vertical Tubes," by C. G. Kirkbride, Trans. AIChE, vol. 30, 1933–34, p. 170.

9 "Characteristics of Flow in Falling Liquid Films," by A. E.

Dukler and O. P. Bergelin, Chemical Engineering Progress, vol. 48, 1952, pp. 557–563.

10 "Numerical Methods in Engineering," by M. G. Salvadori and M. L. Baron, Prentice-Hall, Inc., New York, N. Y., 1952, pp. 57–60.

APPENDIX

Solution of Differential Equation

The differential equation, Equation [15], was solved numerically in the following manner: The developed cylinder surface is divided into equal rectangular meshes of width H in the X-direction and width K in the ϕ-direction as shown in Fig. 14. By finite differences, the partial derivatives at the point (X_i, ϕ_j) are given by backward differences as[10]
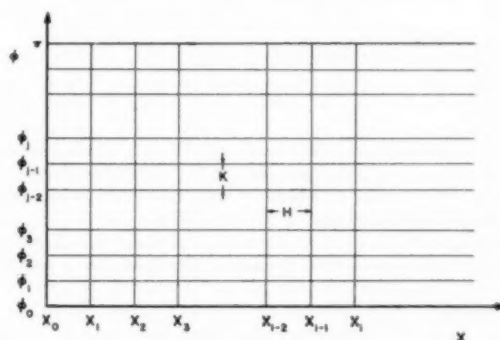
Fig. 14   Mesh construction for numerical solution of Equation [15]

$$\frac{\partial Z_{i,j}}{\partial X} = \frac{1}{2H}(3Z_{i,j} - 4Z_{i-1,j} + Z_{i-2,j})\ldots\ldots[26]$$

and

$$\frac{\partial Z_{i,j}}{\partial \phi} = \frac{1}{2K}(3Z_{i,j} - 4Z_{i,j-1} + Z_{i,j-2})\ldots\ldots[27]$$

where the subscript indexes i, j are employed to denote the point (X_i, ϕ_j) on the mesh.

Substituting from Equations [26] and [27] in Equation [15], one obtains

$$Z_{i,j} = \frac{4}{3A_{i,j}} + \frac{A_{i-1,j}}{A_{i,j}} Z_{i-1,j} - \frac{A_{i-2,j}}{A_{i,j}} Z_{i-2,j}$$
$$+ \frac{A_{i,j-1}}{A_{i,j}} Z_{i,j-1} - \frac{A_{i,j-2}}{A_{i,j}} Z_{i,j-2}\ldots[28]$$

where

$$\left.\begin{aligned}
A_{i,j} &= \frac{3}{2H} + \frac{3\sin\phi_j}{2K} + \frac{4}{3}\cos\phi_j \\
A_{i-1,j} &= 2/H \\
A_{i-2,j} &= 1/2H \\
A_{i,j-1} &= 2\sin\phi_j/K \\
A_{i,j-2} &= \sin\phi_j/2K
\end{aligned}\right\}\ldots\ldots[29]$$

Since the values of the A's in Equation [29] can be evaluated at any point, the value of $Z_{i,j}$ in Equation [28] can be determined at this point if the values of $Z_{i-1,j}$, $Z_{i-2,j}$, $Z_{i,j-1}$, and $Z_{i,j-2}$ are known.

[10] See, for example, Salvadori and Baron (10).

Table 1   Results of numerical solution of Equation [15]

ϕ	0	0.1	0.2	0.3	0.4	0.6	0.8	1.0	1.2	1.6	2.0	2.4	2.8	3.2	∞
0	0	0.124	0.234	0.330	0.413	0.551	0.656	0.736	0.798	0.882	0.931	0.959	0.976	0.987	1.000
15	0	0.125	0.235	0.332	0.417	0.557	0.664	0.746	0.809	0.894	0.945	0.972	0.989	0.999	1.014
30	0	0.126	0.238	0.338	0.425	0.571	0.684	0.771	0.838	0.929	0.982	1.014	1.031	1.042	1.057
45	0	0.128	0.243	0.346	0.439	0.596	0.719	0.815	0.889	0.991	1.050	1.085	1.106	1.118	1.134
60	0	0.129	0.249	0.359	0.459	0.631	0.771	0.881	0.967	1.086	1.157	1.198	1.223	1.237	1.257
75	0	0.131	0.256	0.374	0.484	0.679	0.842	0.974	1.081	1.227	1.316	1.368	1.398	1.416	1.442
90	0	0.133	0.264	0.392	0.515	0.742	0.940	1.106	1.242	1.437	1.555	1.625	1.667	1.691	1.725
105	0	0.135	0.273	0.412	0.551	0.819	1.067	1.285	1.470	1.746	1.919	2.023	2.086	2.122	2.171
120	0	0.137	0.282	0.434	0.591	0.912	1.229	1.528	1.796	2.221	2.503	2.677	2.781	2.842	2.922
135	0	0.139	0.291	0.455	0.632	1.014	1.425	1.846	2.256	2.973	3.499	3.843	4.053	4.179	4.341
150	0	0.140	0.298	0.474	0.669	1.117	1.642	2.236	2.881	4.205	5.376	6.250	6.829	7.186	7.635
160	0	0.141	0.302	0.483	0.688	1.175	1.778	2.507	3.364	5.455	7.629	9.580	11.104	12.123	13.397
170	0	0.142	0.304	0.490	0.701	1.215	1.877	2.724	3.795	6.760	10.977	16.256	21.711	26.624	34.600
180	0	0.142	0.305	0.492	0.706	1.231	1.917	2.815	3.989	7.531	13.586	23.938	41.918	73.539	∞
180	0	0.142	0.306	0.492	0.704	1.226	1.907	2.792	3.953	7.440	13.397	23.533	40.804	70.283	∞
Error, per cent	...	...	...	...	0.31	0.43	0.54	0.81	0.91	1.23	1.41	1.72	2.73	4.63	...

The value of Z along the line $X = 0$ is known to be zero from the boundary condition, Equation [16]. Further, the values of Z along $\phi = 0$ and $\phi = \pi$ can be determined from the second boundary condition, Equation [17], as follows:

Placing $\phi = 0$ into Equation [15] and observing that $(\partial Z/\partial \phi)_{\phi=0} = 0$, the differential equation, Equation [15], is reduced to

$$(\partial Z/\partial X)_{\phi=0} = \frac{4}{3}(1 - Z) \dots \dots [30]$$

The left-hand member of this equation may now be regarded as a total derivative. Using the first boundary condition, Equation [16], Equation [30] would be

$$Z_{\phi=0} = 1 - e^{-\frac{4}{3}X} \dots \dots [31]$$

This gives the value of Z along $\phi = 0$ and shows that $Z_{\phi=0} = 1$ at $X = \infty$.

In a similar manner, for $\phi = \pi$

$$Z_{\phi=\pi} = e^{\frac{4}{3}X} - 1 \dots \dots [32]$$

This shows that $Z_{\phi=\pi}$ tends to infinity as X goes to infinity as was mentioned previously.

For the computations, the mesh sizes were chosen as follows:

$$H = 0.05 \quad \text{for} \quad 0 \leq X \leq 0.2$$
$$= 0.10 \quad \text{for} \quad 0.2 < X \leq 3.2$$
$$K = 2^1/_2^\circ \quad \text{for} \quad 0^\circ \leq \phi \leq 170^\circ$$
$$= 1^1/_4^\circ \quad \text{for} \quad 170^\circ < \phi \leq 180^\circ$$

To use Equation [28] it is necessary to evaluate the function along $X = 0.05$ and along $\phi = 2^1/_2$ deg. In view of the small difference between $Z(0.05, 0)$ and $Z(0.05, \pi)$ as determined from Equations [31] and [32], the value of Z was assumed to change linearly along $X = 0.05$, hence

$$Z(0.05, \pi) = Z(0.05, 0) + \frac{\phi}{\pi}[Z(0.05, \pi) - Z(0.05, 0)] \dots [33]$$

Also, in view of the small difference between $Z(\infty, 0^\circ)$ and $Z(\infty, 2^1/_2^\circ)$, which could be determined from Abramowitz's table (3), the value of Z along $\phi = 2^1/_2$ deg was taken from

$$Z(X, 2^1/_2^\circ) = \frac{Z(\infty, 2^1/_2^\circ)}{Z(\infty, 0^\circ)} Z(X, 0^\circ) \dots \dots [34]$$

Equations [16, 31, 33, 34] give the values of $Z_{0.j}$, $Z_{1.j}$, $Z_{i.0}$, and $Z_{i.1}$, respectively. With these values known, the value of $Z_{1.2}$ can be determined by substituting $H = 0.06$, $K = 2.5\,\pi/180$, and $\phi = 5$ deg into Equation [28] together with the values of $Z_{1.1}$, $Z_{1.0}$, $Z_{1.2.2}$, and $Z_{0.2}$ determined previously.

Employing the functional values already calculated, the value of $Z_{1-1/2.2}$ can be calculated in a similar manner. Step by step, the values of the function at successive points of the mesh can be determined; taking into consideration, of course, the changes in the mesh size, if any.

The values of the function on the element $\phi = \pi$ were calculated numerically from the values of the function at the adjacent points using, instead of Equation [27], a similar expression giving $\partial Z_{i.j}/\partial \phi$ in terms of forward differences (10).

The accuracy of the numerical calculations can be checked by comparing the values of the function at $\phi = \pi$ obtained numerically with those obtained from Equation [32].

The results of the numerical calculations are condensed in Table 1. Originally the calculations were carried out to five decimal places (4). The last two rows of figures in the table are, respectively, the exact values of Z at $\phi = \pi$ as calculated from Equation [32] and the percentage deviation of the numerically calculated values from the exact values. The error in the numerical calculations for $\phi = \pi$ increases slowly to about 2 per cent at $X = 2.6$ after which it rises rather rapidly to 4.63 per cent at $X = 3.2$. This error, for a given value of X, increases slowly except near $\phi = \pi$ to reach the values given in Table 1.

The Third Law of Thermodynamics: A Half-Century Appraisal of the Nernst Heat Theorem

By J. H. POTTER[1], HOBOKEN, N. J.

Fifty years have passed since Nernst formulated his "New Heat Theorem," which he subsequently referred to as the "Third Law of Thermodynamics." The several statements of the theorem are examined in the light of classical thermodynamics, quantum considerations, and recent low-temperature experiments. The limitations under which the heat theorem may be considered a basic scientific law are cited.

Nomenclature

THE following nomenclature is used in the paper:

C = degrees, centigrade
C_x = heat capacity, general
C_p = heat capacity at constant pressure
C_v = heat capacity at constant volume
d = inexact differential
F = Helmholtz function, $U - TS$
ΔF = Helmholtz free energy, $\Delta U - T(\Delta S)$
G = Gibbs function, $H - TS$
ΔG = Gibbs free energy, $\Delta H - T(\Delta S)$
H = enthalpy
J = mechanical equivalent of heat
K = chemical equilibrium constant; also degrees Kelvin
ΔL = isothermal energy change as in a phase transformation
P = pressure
Q = heat
R = gas constant
S = entropy
T = temperature
U = internal energy
V = volume
W = work
γ = Sommerfeld specific heat
θ = thermodynamic temperature
Θ = Debye temperature

Preface

Modern process engineering operates against a backdrop of applied science developed largely from fundamental thermodynamics. At the present time there is great interest in low temperature technology, and much speculation in the area of cryogenics. In this sphere the Third Law of Thermodynamics is of paramount importance.

Object

For half a century the Third Law has occupied a more or less

controversial position in the literature of thermodynamics. Initially widely accepted, the Third Law subsequently underwent a long period of challenge, then gradually won an established place among the fundamental laws of science.

At the time of the formulation of the Third Law, many chemists believed that a logical system of heat calculations should be based upon absolute zero temperature and particularly that entropy should be reckoned above this datum. However, the establishment of a universal origin for entropy was not of prime importance to mechanical engineers, for in conventional heat-to-work conversions the entropy changes rather than the absolute entropy values were sought. Practicing engineers were accustomed to the use of tables of thermodynamic properties of heat mediums, in which zero entropy values had been fixed arbitrarily. For steam (1)[2] zero entropy was set for saturated liquid at 32 F; for ammonia (2) entropy values were measured from saturated liquid at (-40 F). For each of these familiar substances the entropy datum could have been fixed arbitrarily at any other temperature.

The chemists brought a new light to bear upon the nature of entropy calculations, seeking (a) an entropy datum based upon physical considerations, and (b) an understanding of the attainability of absolute zero temperature. Each of these ideas is reflected in a separate statement of the Third Law of Thermodynamics, and each has had both experimental and theoretical justification.

The object of the research upon which this paper was based was to assemble and analyze the several statements of the Nernst heat theorem and, in the light of modern experimental evidence, to determine under what conditions the heat theorem can be considered a basic thermodynamic law.

Introduction

It is a commonplace to state that the First and Second Laws of thermodynamics rest not upon specific conclusive experiments, but rather upon the accumulated experience of the ages and the negative results of innumerable experiments. Bridgman (3) has commented upon the sweeping universality of these laws and of the "uncomfortable" feelings that they stir in some physicists.

The First Law, deduced after the cannon-boring experiments of Rumford (4), the ice block demonstration of Davy (5), and the classic work of Joule (6), asserts that heat and work are mutually interconvertible. The equation

$$W/J = Q \dots \dots \dots \dots \dots \dots [1]$$

represents the First Law as it was originally revealed in the early experiments. However, this is a restricted form of the broader concept of the conservation of energy, postulated by many of the pioneers of thermodynamics, and stated admirably by Carnot (7). In some of the more modern examinations of the First Law (8, 9) emphasis has been placed upon the cyclic behavior of $\oint dQ$

and dW. As $J dQ - dW]$ is a thermodynamic property, then

$$J \oint dQ = \oint dW \dots \dots \dots \dots \dots [2]$$

The Second Law has been subject to many interpretations and has been cast in a wide range of statements. One of the best known is that identified as the Kelvin-Planck (10) formulation, which states that an engine cannot produce net work in a cyclic manner if it exchanges heat with only one isothermal reservoir. This is equivalent to saying that there must be a sink as well as a source, with the clear implication that there must be a limitation on the conversion of heat to work.

Important corollaries developed from the Second Law include the concept of entropy and the statement that no engine operating between two isothermal heat reservoirs can exceed the efficiency of a reversible engine operating between the same two reservoirs. The Carnot cycle, bounded by two isothermal paths and two isentropic paths, has the particular advantage that the only heat added to the cycle is that along the upper isotherm, and the only heat rejected is that along the lower isotherm.

Kelvin (11) appreciated the fact that any property dependent solely upon temperature could be used to define a temperature scale, and proceeded to establish a thermodynamic scale based upon the characteristics of Carnot cycles. As a definition, let

$$\frac{\theta_1}{\theta_2} \equiv \frac{Q_1}{Q_2} \dots \dots \dots \dots \dots [3]$$

where θ_1 and θ_2 are temperatures on a thermodynamic scale and the values of Q correspond to heat addition and rejection in a Carnot cycle. It is to be noted that although Equation [3] defines a temperature scale, the magnitude of the individual degree has not been fixed.

In Fig. 1 three Carnot cycles have been shown connected in cascade. The heat rejected from cycle I becomes the heat added in cycle II, which in turn rejects the heat absorbed in the third cycle. The successive rejection temperatures become

$$\theta_2 = \theta_1 \left(\frac{Q_2}{Q_1} \right); \quad \theta_3 = \theta_2 \left(\frac{Q_3}{Q_2} \right); \quad \theta_4 = \theta_3 \left(\frac{Q_4}{Q_3} \right)$$

For a given heat addition along the path a-b, the over-all cycle efficiency is increased as the sink temperature is progressively lowered. This is equivalent to the statement that the total work of the cascaded system of cycles increases as the sink temperature is lowered. It will be shown that in the limiting value of θ, the summation of the work produced in all the cycles approaches the original heat addition along a-b.

The efficiency of a Carnot cycle is independent of the working substance, for the necessary and sufficient conditions are only that the system be closed and that the successive reversible paths be followed. Traditionally the classroom presentation of the Carnot cycle has been linked to the gas laws for ease in calculations. If a "perfect" gas is used in a Carnot cycle, the thermodynamic temperature may be related to this medium. For a perfect gas, having constant specific heat confined in a vessel of constant volume, the pressure will vary linearly with the temperature. This is the basis for the constant-volume gas thermometer.

In Fig. 2 the pressure-temperature plot is shown for a constant-volume gas thermometer. The pressures at 1 and 2 may be those of fixed points such as the melting or boiling points of particular substances. The temperatures t_1 and t_2 are defined arbitrarily in terms of the behavior of the thermodynamic medium.

Linear extrapolation to an imaginary "zero" pressure will establish an origin for temperatures $0'$ above which the two points will have new numerical values T_1 and T_2. Such a scale is identified as the gas-thermometer scale. It is identical with the
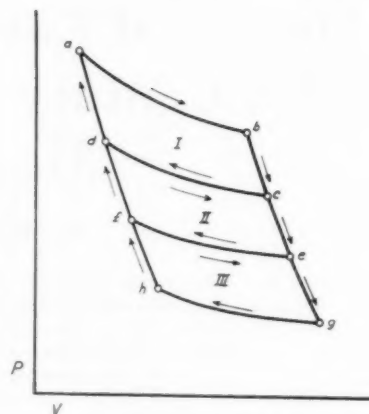

Fig. 1 Carnot cycles connected in cascade
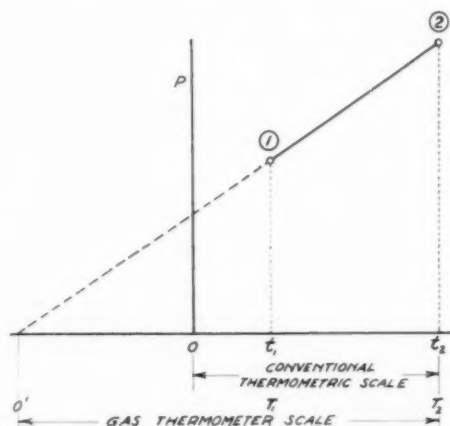
Fig. 2 P-T plot for constant volume gas thermometer

thermodynamic scale if the medium is a perfect gas. Elaborate correction procedures have been developed (12) for use with real gases, to make gas thermometry a practical technique. The zero temperature realized by linear extrapolation of the ideal gas thermometer is a value approached, but never quite attained, by cascading an infinite number of Carnot cycles.

This last statement may be explained with the aid of Fig. 3. This may be considered to be the nth of a cascaded series of Carnot cycles. The lower temperature has been taken as absolute zero, the source at only a very small ΔT higher. The cycle has been drawn in the conventional manner without regard to the actual paths that might be needed for a real heat medium. As the sink temperature is zero, and the temperature at the end of the expansion b-c is also zero, no heat is rejected from the cycle. Therefore all of the heat supplied along a-b must be converted into useful work. Such a condition contravenes the Second Law, and therefore absolute zero temperature is unattainable from a cascade of Carnot cycles. In order to have correspondence between the ideal gas-thermometer scale and the thermodynamic scale at every temperature, one author (13) has suggested that the zero on the Kelvin scale be defined in terms of an ideal reversible cycle with an efficiency of 100 per cent. However, this does not upset the viewpoint that the absolute zero temperature is unattainable.[3]

___

[3] There is not complete agreement on this point, see Appendix.
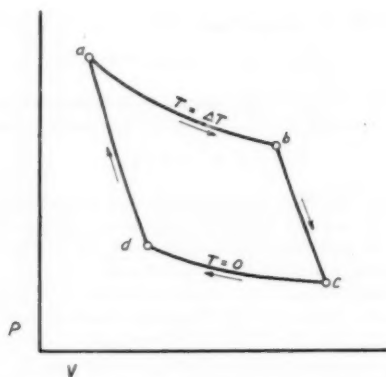
Fig. 3   Ultimate Carnot cycle operating between $T = \Delta T$ and $T = 0$
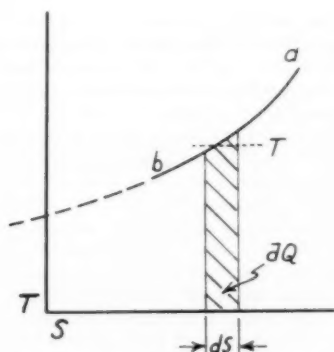
Fig. 4   Small reversible heat abstraction in T-S plane

The entropy concept was mentioned as another significant corollary of the Second Law. Considering only the thermodynamic approach, for a closed system, the entropy change is defined by

$$dS \equiv \left(\frac{dQ}{T}\right)_{rev} \quad\quad\quad\quad [4]$$

While it is not possible to form a physical picture of entropy, this quantity may be represented as a co-ordinate in a plane. In Fig. 4 a reversible path, a-b, is shown between two temperatures T_a and T_b. A small amount of heat dQ is abstracted along this path at an average temperature T. If a linear scale is chosen proportional to the temperature, and an area scale is chosen proportional to the heat abstracted dQ, then the base of the shaded area satisfies Equation [4] and represents dS.

Integration of Equation [4] brings to light a new problem. In general

$$\int_1^2 dS = (S_2 - S_1) = \int_1^2 \left(\frac{dQ}{T}\right)_{rev} = \int_{T_1}^{T_2} C_z \frac{dT}{T} \ldots [5]$$

where C_z is a specific heat capacity that may be a function of T or of both T and P. In the extremely low-temperature ranges the working substance may be in the solid state, where $C_p \cong C_v$. In any event, integration of Equation [5] from zero temperature to temperature T, yields

$$S_T = \int_0^T C_z \frac{dT}{T} + S_0 \ldots\ldots\ldots\ldots [6]$$

where S_0 is a constant of integration. This corollary of the Second Law has made available a new property, entropy, without a datum value for it.

To recapitulate: The First Law established the interconvertibility of heat and work; the Second Law set the limitation on the convertibility of heat to work in terms of reversible engines operating between isothermal reservoirs; the Second Law made possible the thermodynamic temperature scale and introduced the concept of entropy.

The Third Law

The Third Law of thermodynamics grew of the chemists' quests for general and specific information on gas reactions. The basic relationship underlying the Third Law is the Gibbs-Helmholtz equation

$$\Delta F - \Delta U = T \left(\frac{d(\Delta F)}{dT}\right) \ldots\ldots\ldots\ldots [7]$$

Partington (14) has pointed out that this equation was derived originally by Lord Kelvin (15) in 1855, who subsequently offered an alternative proof some 43 years later (16). From Equation [7] the important reaction isochore equation was developed

$$\left.\begin{aligned} \frac{d \ln K_v}{dT} &= \frac{\Delta U}{RT^2} \\[1em] \frac{d \ln K_p}{dT} &= \frac{\Delta H}{RT^2} \end{aligned}\right\} \quad\quad\quad\quad [8]$$

This relationship, generally identified as the van't Hoff equation, is one of the prime tools of the chemist. In view of the physical data available at the time, it was found that the constant of integration for Equation [8] could not be evaluated. The complex nature of this constant did not yield to either the First or Second Laws of thermodynamics. Le Chatelier (17) fully appreciated the importance of the problem as early as 1888; Nernst (18) undertook an extensive investigation which led him back to Equation [7] and through it to the Third Law.

On some experiments with galvanic cells, T. W. Richards (19, 20) noted that the heat of reaction and the electrical energy approached each other in value as the temperature decreased. He reported curves of the types shown in Fig. 5; (a) where the values were equal, (b) where the extrapolated junction would occur well above zero temperature, and (c) where the extended curves met only at zero.
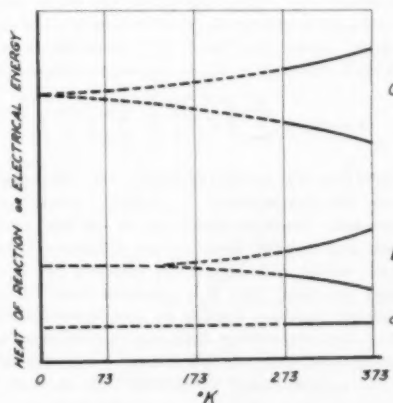

Fig. 5   Representative curves of heat of reaction and electrical energy as shown by Richards

In Fig. 5 it is to be noted that the extrapolations were very liberal, and that in each case the limiting value was reached without passing through a maximum or a minimum. It can be shown that the variables in Fig. 5 could be replaced by ΔG and ΔH, or for that matter by ΔF and ΔU. At least one author (22) contends that Richards' graphs furnished the clues which led Nernst to the formulation of the third law. However, Nernst had shown 8 years previously (23) that $\Delta F \rightarrow \Delta U$ for concentrated solutions of sulfuric acid. Nernst (24) worked generally from Equation [7] and developed his theorem around it. While Richards' graphs imply the generalizations embodied in the Third Law (27) the credit for the original formulation is generally conceded to Nernst

In Fig. 6 a group of characteristic plots similar to those indicated by Richards, has been shown. Each set of curves represents a series of isothermal reversible processes for a condensed system. For any one of these plots it might be claimed that

$$\lim_{T \to 0} (\Delta G - \Delta H) = 0 \dots\dots\dots [9]$$

For curves (a) and (d) an additional statement is possible

$$\lim_{T \to 0} \left(\frac{\partial \Delta G}{\partial T} \right)_p = \lim_{T \to 0} \Delta S = 0 \dots\dots [10]$$

This is not true for curves (b) or (c); in the former ΔG passes through a maximum before reaching the temperature limit; in the latter the ratio $(\partial \Delta G / \partial T)_p$ is still finite at zero temperature. Equation [10] constitutes one formulation for the Third Law; namely, that ΔS approaches zero as T approaches zero for isothermal reversible changes in a condensed system. However, it is difficult to accept an extrapolation of data taken at and above room temperature as a basis for a theorem valid in the region of absolute zero.

The early statements and conjectures about the Third Law produced several significant effects. These were (a) to encourage research on physical constants, particularly specific heats, (b) to stimulate theoretical studies, especially quantum mechanics, and (c) to spur cryogenic investigations. Each of these areas will be touched on briefly.

Investigations into the behavior of specific heats disclosed some very complex phenomena, which in turn caused the entropy equations to become more complicated. An example in point is that of nitrogen (55) for which a part of the T-S diagram is shown in Fig. 7. Condensation takes place along g-f and solidification from e to d. A solid-phase transformation takes place along c-b, and each of these steps is connected by a path for which C_p is some different function of T. The entropy at any point would then include the zero point entropy S_0, plus the sum of all of the n temperature-related paths, plus the sum of all of the m isothermal changes. This is indicated in the generalized entropy equation

$$S = S_0 + \sum^n \int C_p \frac{dT}{T} + \sum^m \left(\frac{\Delta L}{T} \right) \dots\dots [11]$$

As the Third Law is concerned primarily with events which take place at very low temperature, the entropy investigation also may be restricted. Most substances are in the solid state at low temperatures and exhibit little or no difference between C_p and C_v. One trivial case should be disposed of: If the heat capacity were constant, then it is apparent from Equation [6] that the entropy becomes infinite at zero temperature. Both theory and experiment confirm that heat capacities are temperature-dependent, even at low temperatures. As one of the interesting peripheral contributions, Gruneisen (32) showed that for metals the ratio of the coefficient of expansion to C_p was constant at all temperatures. While this again was an approximation, it suggested the dependence of C_p on T.

At extremely low temperatures, Debye (33, 34) showed that

$$C_p \approx C_v = 464.4(T)^3/\Theta^3 \dots\dots\dots [12]$$

where Θ is the "Debye temperature," a characteristic of a given substance. For any particular material this could be restated as

$$C_p \approx C_v = k(T)^3 \dots\dots\dots [13]$$

Inspection of Equation [13] indicates that the integral term in Equation [6] does not become indeterminate at very low temperatures. A graphical demonstration is given in Fig. 8. Here C_p/T has been plotted against T in the range from 0 to 1 deg K. It is apparent that the solid line representing C_p/T goes to zero faster than the broken line representing a linear decrease in T.

More recent investigations of specific-heat capacities at low temperatures indicate that for metals a linear term in T must be added to the Debye equation to account for the specific heats of the electrons (35, 36). However, this does not invalidate Equation [6].

Simon (37) has presented several interesting charts which tend to show both aspects of the Third Law. These have been adapted in the next four figures. In Fig. 9 an imaginary engine is caused to operate between pressures P_1 and P_2 for a working substance which has constant specific heats. As cited previously, this is a trivial case for which the entropy tends to infinity as $T \rightarrow 0$. Cooling may be realized by isothermal compression ab, followed by isentropic expansion bc. However, no number of such engines could reduce the temperature to zero.

Einstein's simple theory of 1907 (38), accounting for the decrease in specific heats of solids with decreasing temperature, was among the first important triumphs of quantum mechanics. This theory predicted that the specific heat tended to zero as the temperature approached zero. If in addition it is assumed that C_p/T approaches zero faster than T approaches zero, there is the possibility that a set of curves such as those shown in Fig. 10 might exist. Here each parameter V has a different value of entropy at $T = 0$. As a result, a multistaging of isothermal and isentropic paths would lead to absolute zero temperature in a finite number of steps. It will be seen presently that the operations suggested by Fig. 10 cannot be realized.

Planck (39) maintained that for pure crystalline substances the actual value of the entropy at zero temperature was zero. This is indicated in Fig. 11 when the parameters V meet at zero. Here the multistage cooling operation fails to reach zero temperature regardless of the number of stages employed. Modern workers in this field (40) challenge the statement that at absolute zero the entropy is also zero, on both statistical and conventional approaches to the entropy concept. For glasses and reactions in which a false or "frozen" equilibrium exists there will be a residual entropy; however, Equation [10] will be valid in that there will be no entropy difference between the coexisting states. The generalized T-S diagram would then be that shown in Fig. 12, with all the parameters intersecting at absolute zero temperature but with a finite entropy S_0.

It was mentioned previously that the coefficient of expansion was an important index to the temperature effect on the specific heat. An independent check method was suggested by Lindemann (41). The coefficient of expansion is given by

$$\beta = \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_p$$

Also the Maxwell relation

$$\left(\frac{\partial S}{\partial P} \right)_T = -\left(\frac{\partial V}{\partial T} \right)_p \dots\dots\dots [14]$$
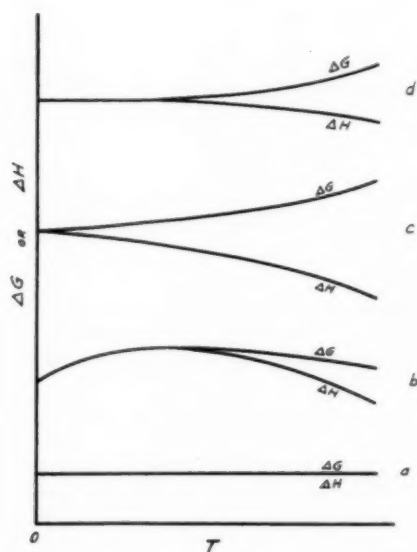
Fig. 6   Typical plots of enthalpy change and Gibbs free-energy change at low temperatures

Fig. 7   Temperature-entropy plot for nitrogen

Fig. 8   Plot of C_p/T versus T for a Debye substance
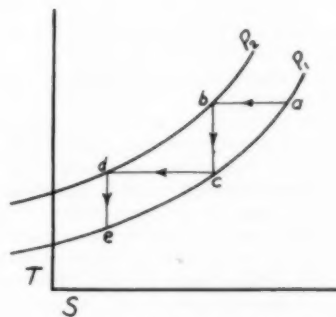
Fig. 9   Multistage cooling for a material having constant specific-heat
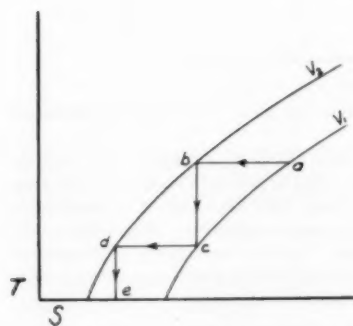
Fig. 10   Multistage cooling for a material with $C_v = f(T)$
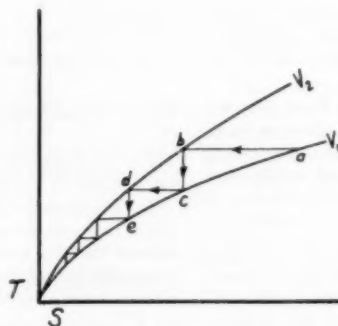
Fig. 11   Multistage cooling for a material with $C_v = 0$ at $T = 0$, $S = 0$
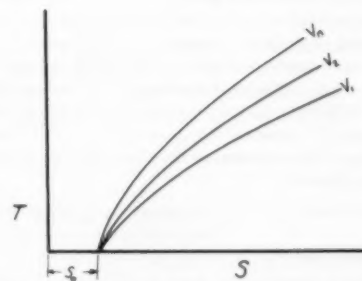
Fig. 12 T-S diagram for a substance having finite entropy at $T = 0$

Theory requires that

$$\lim_{T \to 0}\left(\frac{\partial S}{\partial P}\right)_T \to 0$$

By experiment on a number of metallic substances it was shown that

$$\lim_{T \to 0}\left(\frac{\partial V}{\partial T}\right)_p \to 0$$

Buffington (42) found that quartz glass and supercooled silicon did not meet the criterion cited at low temperatures. This suggested that Planck's requirement of a "pure crystalline substance" was indeed a necessary condition in the statement of the Third Law.

The wide publicity given the magnetic approach to absolute zero temperature precludes any need to recount the details of the process here (52). It was suggested independently by Debye (53) and Giauque (54) at about the same time. A paramagnetic salt is cooled in a cryostat with liquid helium. The steps in the process are indicated in Fig. 13. At the lowest precooling temperature, T_a, a magnetic field is applied. During magnetization the salt is maintained at constant temperature. When the magnetic field is withdrawn, the salt undergoes isentropic cooling to a temperature T_c. It has been suggested (36) that magnetic cooling might be applied in a cascaded manner, as shown in Fig. 14. Actually this constitutes an independent demonstration of the unattainability statement of the Third Law, and has significance whether or not the experimental work can be realized.

An interesting substantiation of the Third Law involves the determination of the entropy of a crystalline compound which can exist in more than one form. In essence the entropy change is measured between two temperatures in each of two allotropic forms of the material. Stephenson and Giauque (43) worked on two forms of phosphine between 0 and 59.43 K, finding in one case an entropy of 8.14 cal/deg mol and in the other 8.13 cal/deg mol.

Darken and Gurry (44) made an interesting analysis of the allotropic forms of tin. An imaginary cycle is set up as shown in Fig. 15. Gray tin is stable below 19 C, and it is assumed that gray tin is cooled to absolute zero to start the cycle at a. Heat is then applied to bring the tin to the transition temperature b. From b to c there is an isothermal transition from gray tin to white tin. By rapid cooling the white tin is cooled to absolute zero along c-d. In the last imaginary step, the tin changes over to the gray form at absolute zero. The sum of the entropies in the first three paths is almost equal to zero, as shown in Fig. 16. The writers were inclined to believe that the small entropy difference between d and a was of the order of magnitude of the experimental error, and that the entropy difference between the two forms of tin at zero temperature was zero.

An interesting experimental verification of the Third Law was found in a cryogenic investigation of helium (56). Even as low as 1 K, helium may exist in vapor, liquid, or solid phase depending upon the pressure. Simon and Swenson (57) applied the Clausius-Clapeyron equation to the phase change between liquid helium II and the solid

$$dP/dT = \Delta S/\Delta V \dots\dots\dots\dots [15]$$

if the Lim

$$\lim_{T \to 0}(dP/dT) \to 0$$

and as ΔV remains finite, then

$$\lim_{T \to 0}\Delta S \to 0$$
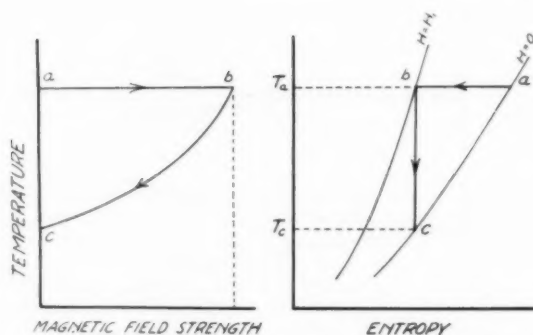
Fig. 13   Steps in cooling technique using a paramagnetic salt
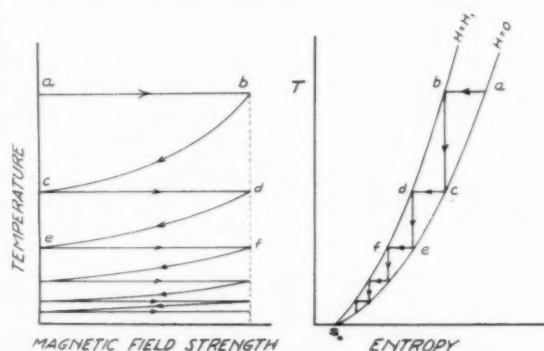
Fig. 14   Proposed multistage cooling technique using paramagnetic salt
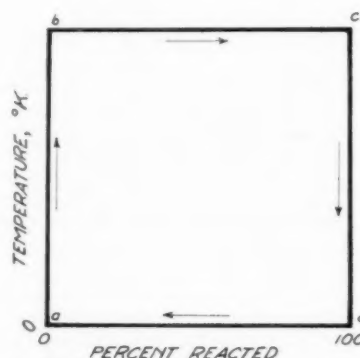
Fig. 15   Temperature-composition chart for allotropic transformation

The experiments showed that $(dP/dT) = 0.425\,(T)^7$, so that for small values of T, (dP/dT) did in fact tend rapidly toward zero.

Other cryogenic phenomena also have helped to explain the Third Law. In his work on superconductivity, Mendelssohn (37) showed that for a number of metals the entropy difference between the normal and superconducting states approached zero as a linear function of the absolute temperature

$$(S_n - S_0) = -\gamma T$$

$$\lim_{T \to 0}(S_n - S_0) \to 0 \dots\dots\dots\dots [16]$$

Hence the entropy change, even in this unusual state, tends to zero as the temperature approaches zero.
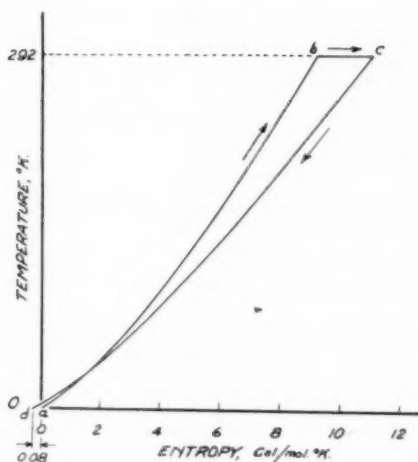
Fig. 16 Temperature-entropy diagram for allotropic transformation cycle for gray and white tin

Workers in the field of statistical mechanics have overcome earlier doubts (48) as to the validity of the Third Law, as more of the experimental evidence has been produced. As Simon (71) stated, there is great plausibility in the assumption that the state of lowest energy is also that of lowest entropy, or of perfect order. Recently there has been a widespread interest in the statistical mechanics of mixtures of (62, 63, 64, 65) isotopes in the low temperature areas.

A new and rather startling concept has come about as a result of studies in nuclear spin systems (68, 69, 70). It is the concept of "negative" temperature. Simon has indicated that the existence of negative temperatures will not necessarily prevent operations subject to the ordinary thermodynamic laws. In Fig. 17 some conjectural plots have been drawn showing the conventional T-S plot at (a), the $1/T$-S plot in the region of $T = 0$ and the $1/T$-S plot in the region of $T = +\infty$. The unattainability statement of the Third Law would still apply, with the additional proviso that $T = 0$ could be approached from a positive or negative direction, Fig. 17(b).

Discussion

The Third Law has had a stormy half century during which it has undergone certain modifications owing to theoretical and experimental advances in the areas of physics and chemistry. Across the years Simon was the uncompromising champion of the Third Law, which he considered . . . "the greatest advance in thermodynamics since van der Waals' time" (71).

The contention that the Third Law should be restricted to crystalline substances had to be overcome, as even Nernst realized that such a limitation would preclude the acceptance of the theorem as a basic scientific law. In his 1906 presentation Nernst specified that for condensed systems

$$\text{Lim } (\Delta S) \to 0$$
$$T \to 0$$

The theorem extended to cover gases in 1908; the unattainability statement came in 1910 (25).

As was shown in the Introduction, the unattainability of absolute zero temperature can be inferred as a consequence of the Second Law. However, the unattainability proofs afforded by the low-temperature characteristics of the specific heats, and the multistaging of a paramagnetic material are unique and are not dependent upon the Second Law.

The thermodynamic tests of the Nernst Heat Theorem, including the coefficient-of-expansion method and the cyclic comparison of allotropic transformations, tend to establish it as a scientific law. While specific-heat data based upon calorimetric and spectroscopic techniques do not always agree, the differences may often be reconciled in terms of experimental error and the limitations of instrumentation. In many cases the two types of measurement of specific heat check well enough to support the Third-Law interpretations of these quantities at very low temperatures.

In the superconductors and in the response to the Clausius-Clapeyron equation of helium II, great new evidences of the validity of the Third Law are presented. Lastly, even the concept of "negative" temperatures is not incompatible with the theorem.

If some of the original objectives of the Third Law, such as the zero point for entropy, are less important today than when the search began, the chain of research and speculation which it provoked lighted the way to important new fields. The progress which this record unfolds follows the classic pattern described so well by Conant (72), in which one set of experiments laid the basis for others with a burgeoning of peripheral theories and techniques.
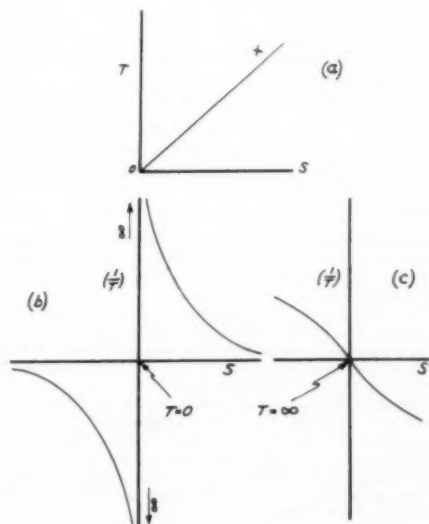
Fig. 17 Negative temperature concept: (a) Conventional T-S plot; (b) plot of $(1/T)$-S near $T = 0$; (c) plot of $(1/T)$-S near $T = \infty$

The majestic figures of Clausius and Kelvin loom large in the background of this entire development. To them are due the concepts of entropy, the Second Law, the thermodynamic temperature scale, and the other relations upon which Nernst was able to fashion the Third Law.

Conclusions

1 The Third Law of thermodynamics has won an established place among the fundamental scientific laws.

2 The unattainability of absolute zero temperature constitutes one of the forms of the Third Law of thermodynamics. Two statements of the unattainability principle are cited below:

"It is impossible to devise an arrangement by which a body may be completely deprived of its heat, i.e., cooled to the absolute zero." (Nernst)

"It is impossible by any procedure, no matter how idealized, to

reduce the entropy of a system to its zero-point value in a finite number of operations." (Zemansky)

3 The Third Law may also be stated in terms of the behavior of entropy. Two recent formulations are given:

"The entropy of all factors within a system which are in internal thermodynamic equilibrium disappears at absolute zero." (Nernst-Simon)

"At zero temperature a system may assume several states having less entropy than all other states." (Keenan)

Acknowledgments

It is a pleasure to acknowledge the friendly assistance of a number of people in the preparation of this paper. Many important references were brought to the author's attention through correspondence with men actively at work in areas related to the Third Law. In this country helpful advice or suggestions were sent me by Messrs. D. H. Andrews, F. G. Brickwedde, S. C. Collins, G. T. Furukawa, W. F. Giauque, F. G. Keyes, and D. R. Stull. Prof. I. Prigogine, at the Free University of Brussels, and Profs. J. F. Allen and K. Mendelssohn in Britain were most helpful. Professor emeritus J. R. Partington, of the University of London, who personally knew both Richards and Nernst, sent suggestions and references for which the author is most grateful.

Thanks are due to Profs. C. F. Kayan, J. H. Keenan, J. R. Partington, L. Z. Pollara, M. W. Zemansky, and to Dr. Warren De Sorbo, each of whom was kind enough to read the manuscript and to give the benefit of his critical opinion.

The second entropy statement of the Third Law was communicated by Prof. J. H. Keenan and has not been published previously. The material in the Appendix was supplied by Prof. M. W. Zemansky and was reproduced with his permission.

Bibliography

1 "Thermodynamic Properties of Steam," by J. H. Keenan and F. G. Keyes, first edition, John Wiley & Sons, Inc., New York, N. Y.

2 "Tables of Thermodynamic Properties of Ammonia," Circular no. 142, National Bureau of Standards.

3 "The Nature of Thermodynamics," by P. W. Bridgman, Harvard University Press, Cambridge, Mass., 1941.

4 "An Inquiry Concerning the Source of the Heat Which is Excited by Friction," by Benjamin Rumford, Philosophical Transactions, (1798) vol. 18, 1796–1800, pp. 278–287.

5 "Collected Works," by Sir Humphry Davy, 1799.

6 "On the Calorific Effects of Magneto-Electricity, and on Mechanical Value of Heat," by J. P. Joule, British Association, 1843.

7 "Réflexions sur la Puissance Motrice du Feu," by N. L. S. Carnot, Chez Bachelier, Libraire, Paris, France, 1824, ASME, 1943.

8 "Thermodynamics," by J. H. Keenan, John Wiley & Sons, Inc., New York, N. Y., 1941.

9 "A New Examination of the Laws of Thermodynamics," by R. B. Green, American Journal of Physics, vol. 22, 1954, pp. 191–193.

10 "Mechanical Engineering Thermodynamics," by D. A. Mooney, Prentice-Hall, Inc., New York, N. Y., 1953.

11 "Account of Carnot's Theory of the Motive Power of Heat," by Sir Wm. Thompson, Trans. Royal Society of Edinburgh, vol. 16, 1849.

12 "Heat and Thermodynamics," by J. K. Roberts, third edition, Blackie and Son Ltd., London, England, 1940.

13 "Thermodynamics for Chemists," by S. Glasstone, first edition, D. Van Nostrand Company, Inc., New York, N. Y., 1947.

14 "The Nernst Memorial Lecture," by J. R. Partington, Journal of the Chemical Society, Sept. 1953, pp. 2853–2872.

15 "On the Thermo-Elastic and Thermo-Magnetic Properties of Matter," by W. Thomson (Lord Kelvin), Quarterly Journal of Mathematics, vol. 1, 1857, pp. 57–77.

16 "On Thermodynamics Founded on Motivity and Energy," by Lord Kelvin, Proceedings of the Royal Society of Edinburgh, vol. 22, 1900, pp. 126–130.

17 "Recherches expérimentales et théorétiques sur les équilibres chimiques," by Le Chatelier, Ann. Mines, vol. 13, 1888, pp. 336–356.

18 "Experimental and Theoretical Applications of Thermodynamics to Chemistry," by Walther Nernst, Silliman Lectures of 1906, Yale University Press, New Haven, Conn., 1908.

19 "Die Bedeutung der Anderung des Atom volums III," by Th.

W. Richards, Zeitschrift für physikalische Chemie, vol. 42, 1902, pp. 129–154.

20 "The Significance of Changing Atomic Volume," by Th. W. Richards, Proceedings, American Academy of Arts and Sciences, vol. 38, 1902, pp. 293–317.

21 "Thermodynamics and the Free Energy of Chemical Substances," by G. N. Lewis and M. Randall, McGraw-Hill Book Company, New York, N. Y., 1923, p. 437.

22 "Chemical Thermodynamics," by I. M. Klotz, Prentice-Hall Inc., New York, N. Y., first edition, 1950.

23 "Über die mit der Vermischung concentrirter Lösungen verbundene Aenderung der freien Energie," by H. W. Nernst, Annalen der Physik, vol. 53, 1894, p. 57.

24 "The New Heat Theorem," by H. W. Nernst, E. P. Dutton & Co., New York, N. Y., 1926.

25 "The Third Law of Thermodynamics," by E. D. Eastman, Chemical Reviews, vol. 18, 1936, pp. 257–273.

26 "The Entropy of the Elements and the Third Law of Thermodynamics," by G. N. Lewis and G. E. Gibson, Journal of the American Chemical Society, vol. 39, 1917, pp. 2554–2581.

27 "The Third Law of Thermodynamics and the Entropy of Solutions and of Liquids," by G. N. Lewis and G. E. Gibson, Journal of the American Chemical Society, vol. 42, 1920, pp. 1529–1533.

28 "The Entropy of a Crystalline Solution of Silver Bromide and Silver Chloride in Relation to the Third Law of Thermodynamics," by E. D. Eastman and R. T. Milner, Journal of Chemical Physics, vol. 1, 1933, pp. 445–456.

29 "The Entropy of Supercooled Liquids at the Absolute Zero," by L. Pauling and R. C. Tolman, Journal of the American Chemical Society, vol. 47, 1925, pp. 2148–2156.

30 "The Third Law of Thermodynamics Evidence From the Specific Heats of Glycerol That the Entropy of a Glass Exceeds That of a Crystal at the Absolute Zero," by G. W. Gibson and W. F. Giauque, Journal of the American Chemical Society, vol. 45, 1923, pp. 93–104.

31 "Entropy Changes at Low Temperatures," by G. E. Gibson, W. M. Latimer, and G. S. Parks, Journal of the American Chemical Society, vol. 42, 1920, pp. 1533–1542.

32 "Über die thermische Ausdehnung und die spezifische Wärme der Metalle," by E. Gruneisen, Annalen der Physik (4), vol. 26, 1908, pp. 211–216.

33 "Zur Theorie des spezifischen Wärmen," by P. Debye, Annalen der Physik, vol. 39, 1912, pp. 789–839.

34 "Entropy Calculations at Low Temperatures," by George Calingaert, Journal Chemical Education, vol. 31, 1954, p. 487.

35 "The Atomic Heats of Palladium, Sodium, and Mercury at Low Temperatures," by G. L. Pickard and F. E. Simon, Proceedings of the Physical Society, vol. 61, part 1, no. 343, July 1948, pp. 1–9.

36 "Heat and Thermodynamics," by M. W. Zemansky, McGraw-Hill Book Co., New York, N. Y., fourth edition, 1957.

37 "Low Temperature Problems, A General Survey," by F. E. Simon, Low Temperature Physics, Pergamon Press Ltd., London, 1952.

38 "Die Planksche Theorie der Strahling and die Theorie der Spezifischen Warme," by A. Einstein, Annalen der Physik, vol. 22, 1907, pp. 180–190.

39 "Treatise on Thermodynamics," by M. Planck, Dover Publication, New York, N. Y., third edition, 1945.

40 "The Principles of Chemical Equilibrium," by K. G. Denbigh, Cambridge University Press, Cambridge, England, first edition, 1955.

41 "Über die Temperaturabhangigkeit des thermischen Ausdehnungskoeffizienten," by C. L. Lindemann, Physikalische Zeitschrift, vol. 12, 1911, pp. 1197–1199.

42 "The Measurement of Coefficients of Expansion at Low Temperatures," by R. M. Buffington and W. M. Latimer, Journal of the American Chemical Society, vol. 48, Sept. 1956, pp. 2305–2319.

43 "A Test of the Third Law of Thermodynamics by Means of Two Crystalline Forms of Phosphine," by C. C. Stephenson and W. F. Giauque, Journal of Chemical Physics, vol. 5, pp. 149–158, 1937.

44 "The Physical Chemistry of Metals," by Darken and Gurry, McGraw-Hill Book Company, Inc., New York, N. Y.

45 "The Specific Heats of Germanium and Gray Tin at Low Temperatures," by R. W. Hill and D. H. Parkinson, Philosophical Magazine, vol. 43, seventh series, No. 338, March 1952, pp. 309–316.

46 "The Mass Effect in the Entropy of Solids and Gases," by W. M. Latimer, Journal of the American Chemical Society, vol. 43, 1921, pp. 818–826.

47 "Statistical Thermodynamics," by R. H. Fowler and E. A. Guggenheim, Cambridge University Press, Cambridge, England, 1952, pp. 219–229.

48 "Statistical Mechanics With Particular Reference to the Vapor

Pressures and Entropies of Crystals," by R. H. Fowler and T. E. Sterne, *Reviews of Modern Physics*, vol. 4, 1932, pp. 635–722.

49 "Thermodynamics of Technical Gas Reactions," by Fritz Haber, A. B. Lamb translation, Longmans, Green, London, 1908.

50 "Man-Made Diamonds," by F. P. Bundy, H. T. Hall, H. M. Strong, and R. H. Wentorf, *Nature*, vol. 176, 1955, pp. 51–55.

51 "Note on the Theory of Magnetic Storms," by F. A. Lindemann, *Philosophical Magazine*, vol. 38, 1919, pp. 669–684.

52 "Experiments at Very Low Temperatures Obtained by the Magnetic Method," by N. Kurti and F. E. Simon, *Proceedings of the Royal Society*, vol. A-149, pp. 152–176.

53 "Einige Bemerkungen zur Magnetisierung bei tiefer Temperatur," by P. Debye, *Annalen Physik*, vol. 81, 1926, pp. 1154–1160.

54 "A Thermodynamic Treatment of Certain Magnetic Effects. A Proposed Method of Producing Temperatures Considerably Below 1° Absolute," by W. F. Giauque, *Journal of American Chemical Society*, vol. 49, 1927, pp. 1864–1870.

55 Bulletin 394, by K. K. Kelley, U. S. Bureau of Mines, U. S. Government Printing Office, Washington, D. C., 1936.

56 "Liquid Helium," by J. F. Allen, Low Temperature Physics, Pergamon Press, Ltd., London, England, 1952.

57 "The Liquid-Solid Transition in Helium Near Absolute Zero," by F. E. Simon and C. A. Swenson, *Nature*, vol. 165, 1950, pp. 829–831.

58 "A λ Anomaly in the Specific Heat of Solid Hydrogen," by R. W. Hill and B. W. A. Ricketson, *Philosophical Magazine*, vol. 45, 1954, pp. 277–282.

59 "Theoretically Interesting Aspects of High Pressure Phenomena," by P. W. Bridgman, *Reviews of Modern Physics*, vol. 7, 1935, pp. 1–33.

60 "The Third Law of Classical Thermodynamics," by P. C. Cross and H. C. Eckstrom, *Journal of Chemical Physics*, vol. 10, 1952, pp. 287–291.

61 "The Temperature Range Below 1° Absolute," by N. Kurti, Low Temperature Physics, Pergamon Press, Ltd., London, England, 1952.

62 "Statistical Mechanics of Mixtures of Isotopes," by G. V. Chester, *Physical Review*, vol. 100, no. 2, Oct. 15, 1955, pp. 446–454.

63 "Effets Isotopiques et Propriétés Thermodynamiques en Phase Condensée-I," by I. Prigogine, R. Bingen, and J. Jeener, *Physica*, vol. 20, 1954, pp. 383–394.

64 "Effets Isotopiques et Propriétés Thermodynamiques en Phase Condensée—II," by I. Prigogine and J. Jeener, *Physica*, vol. 20, 1954, pp. 516–520.

65 "Effets Isotopiques et Propriétés Thermodynamiques en Phase Condensée—III," by I. Prigogine, R. Bingen, and A. Bellemans, *Physica*, vol. 20, 1954, pp. 633–654.

66 "Treatise on Physical Chemistry," by Taylor and Glasstone, D. Van Nostrand Co., New York, N. Y., 1942.

67 "Kinetic Theory of Gases," first edition, by E. H. Kennard, McGraw Hill Book Co., New York, N. Y.

68 "Nuclear Audiofrequency Spectroscopy by Resonant Heating of the Nuclear Spin System," by N. F. Ramsey and R. V. Pound, *Physical Review*, vol. 81, 1951, pp. 278–279.

69 "A Nuclear Spin System at Negative Temperature," by E. M. Purcell and R. V. Pound, *Physical Review*, vol. 81, 1951, pp. 279–280.

70 "The Concept of Temperature Near Absolute Zero," by F. E. Simon, Temperature, Reinhold Publishing Corp., New York, N. Y., 1955.

71 "On the Third Law of Thermodynamics," by F. Simon, *Physica IV*, no. 10, Nov., 1937, pp. 1089–1096.

72 "On Understanding Science," by J. B. Conant, The New American Library, New York, N. Y., 1951, p. 108.

APPENDIX

General agreement is lacking on the statement that the unattainability of absolute zero temperature can be deduced from the Second Law. In a personal communication from Prof. Mark W. Zemansky, he points out that the Second Law does not afford such proof. He avers that the Second Law is only a generalization from experience with heat engines operating between reservoirs provided by nature. An engine with 100 per cent thermal efficiency could be imagined if nature had provided an infinite sink at zero temperature.

Also, the Second Law does not preclude the possibility of a laboratory experiment in which a small amount of material might be reduced to absolute zero. This material might also have zero heat capacity at absolute zero and could not be used as a reservoir for a heat engine, so that the Second Law would still be valid.

The Effect of Conduit Dynamics on Control-Valve Stability[1]

By F. D. EZEKIEL,[2] CAMBRIDGE, MASS.

A method is presented for simulating the dynamic behavior of a frictionless hydraulic conduit with distributed parameters in conjunction with a simple spool-type control valve for analog-computer use. Analytical and experimental results showing the regions of stable and unstable operation of such a system are given, together with suggested stabilizing techniques. The effect of the valve chamber in the over-all analysis is found to be of considerable importance.

Nomenclature

THE following nomenclature is used in this paper:

a = conduit cross-sectional area, sq in.
b = valve viscous-damping coefficient, lb-sec/in.
c = velocity of sound in fluid, ips
C_q = orifice-discharge coefficient
C_u = orifice-velocity coefficient
E = hydraulic-spring stiffness per unit pressure, in.
F = force acting on valve spool, lb
F_e = external force, lb
F_h = hydraulic force, lb
F_s = spring force, lb
f = nondimensional external force
i = subscript to denote initial condition
K = spring constant, lb/in.
l = conduit length, in.
L = valve damping length (positive when fluid flows in direction so as to produce positive damping), in.
m = equivalent spool mass, lb-sec^2/in.
P = valve upstream pressure and valve pressure drop, psi
P_{crit} = valve upstream pressure for maximum steady-state flow, psi
P_r = relative pressure = P/P_{crit}
p = nondimensional pressure = P/P_i
p_s = nondimensional supply pressure = P_s/P_i
Q = volumetric fluid flow rate, cu in./sec
Q_{crit} = maximum steady-state value of Q
Q_r = relative flow = Q/Q_{crit}
q = nondimensional flow = Q/Q_i
q_o = nondimensional flow through valve orifice
q_l = nondimensional flow at valve end of the conduit
r = nondimensional valve stability parameter = P_i/P_r
t = time, sec

$\mathcal{U}_c$ = valve-chamber volume, cu in.
w = valve port width, in.
X = valve displacement or opening, in.
X_0 = valve displacement with zero spring force, in.
X_s = spring displacement, in.
x = nondimensional valve displacement = X/X_i
β = over-all bulk modulus of fluid in conduit, psi
β_c = over-all bulk modulus of fluid in valve chamber, psi
ζ = damping ratio
ζ_L = parameter denoting a measure of valve damping length
θ = jet angle of outgoing fluid at valve orifice measured from spool axis, deg
ρ = fluid density, lb-sec^2/in.^{4}
ϕ = conduit nondimensional impedance = $\dfrac{\sqrt{(\rho\beta)}}{a}\dfrac{Q_i}{P_i}$
ω_n = valve natural frequency $\sqrt{(K/m)}$, rad/sec
ω_l = fundamental natural frequency for a closed conduit of length $l = \dfrac{\pi}{2}\dfrac{c}{l}$, rad/sec

Introduction

With modern developments in the field of hydraulic control, increasingly stringent requirements are being imposed on the design and operation of valves and other components. The signal input to a control valve is usually at low force and energy levels, while the controlled output is at relatively high power levels. As a result, the relatively small unfavorable interaction between flow, pressure, and valve forces may cause instability.

It is significant to note that undesirable oscillations frequently occur at frequencies well above the fundamental natural frequencies of the loads being driven by the valves. In fact, it is usually possible to connect a valve to a test stand without any load (that is, its output ports connected by an unrestricted path), and observe a tendency to oscillate if the moving part of the valve is held in place with just a mechanical spring.

An analysis by Lee (1)[3] of the unsteady flow of fluid through the passages in the moving part of a valve showed that acceleration of the fluid induces a force on the valve. This analysis showed that when the pressures immediately upstream and downstream of the valve are held constant, this force is either in the same or in the opposite direction to a spool velocity depending on the direction of the stream flow through the valve. For example, the direction of fluid flow shown in Figs. 1 and 2 is that which provides a positive damping effect, because the induced force is in an opposite direction to the spool velocity. Reversing the flow direction in Figs. 1 and 2 brings about a negative damping effect. In either case the magnitude of the damping is proportional to the so-called damping length L as shown in Fig. 2, which will be referred to as a positive damping length when the damping is positive, and a negative damping length when the damping is negative.

Although the application of Lee's analysis proves very helpful in many valve designs, there are some cases in which a valve will oscillate when care has been taken to provide a positive damping length.

[3] Numbers in parentheses refer to the Bibliography at the end of the paper.
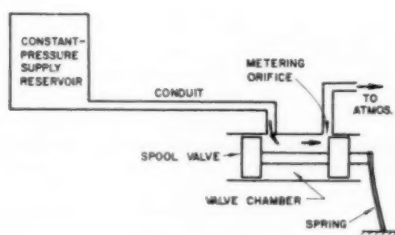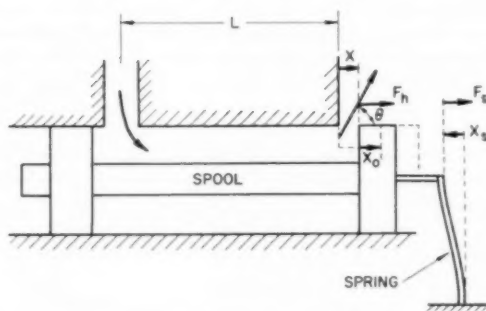
Fig. 1   Valve conduit system

Fig. 2   Spool valve showing steady-state forces

Oscillations With Positive Damping Length

Since a valve having a negative damping length is very unstable, and hence unsatisfactory for control purposes, the following treatment is restricted to the stability study of valves having only positive damping lengths. The arrangement of components shown schematically in Fig. 1 serves to illustrate the system to be analyzed. A simple single-orifice spool valve was connected by means of a supply line to a constant hydraulic-supply pressure source with the outflow from the valve being exhausted directly to the atmosphere. The valve spool was connected to a cantilever spring having a stiffness corresponding to that of an electromagnetic torque motor which might be used to actuate the valve.

The valve, as shown in Fig. 1, was found to be stable under certain operating conditions and unstable under others. For example, varying the length or size of the conduit would bring about greatly varying degrees of valve oscillation. In some instances the valve spool seemed perfectly stable and in others it would oscillate violently. Thus the supply conduit seemed to play an important role in this type of valve oscillation.

It is appropriate to note that some of the concepts discussed here are the same ones that Ainsworth (2) developed independently, but concurrently, at the Minneapolis-Honeywell Regulator Company.

Steady-State Valve Analysis

Fig. 2 represents the spring-loaded spool inside the valve body. The quantity X_0 is the spool displacement when there is no flow or when the spring force is zero. When fluid flows through the valve, a hydraulic force F_A due to the oil momentum (1) is set up

$$F_A = -2C_q C_a \cos \theta \; wPX = -EPX \ldots \ldots \ldots [1]$$

where P is both the upstream pressure and the pressure drop across the valve because the downstream pressure is zero, E is the hydraulic spring stiffness per unit pressure, and X is the valve displacement. The magnitude of E will be considered as constant throughout this analysis, since its variation is fairly small as demonstrated experimentally, Equations [1] and [5].

The spring force F_s shown in Fig. 2 can be defined as in Equation [2]

$$F_s = KX_s = K(X_0 - X) \ldots \ldots \ldots \ldots [2]$$

At equilibrium, the net force acting on the valve must be zero. Hence

$$F_A + F_s = 0 \ldots \ldots \ldots \ldots \ldots [3]$$

The orifice-flow equation yields

$$Q = C_q wX \sqrt{\frac{2P}{\rho}} \ldots \ldots \ldots \ldots \ldots [4]$$

By combining Equations [1] through [4] in order to eliminate X, a relation between Q and P can be established

$$Q = \frac{C_q wKX_0 \sqrt{2} \sqrt{P}}{E \sqrt{\rho} \left(P + \dfrac{K}{E} \right)} \ldots \ldots \ldots \ldots [5]$$

By differentiating Equation [5] with respect to P and setting $dQ/dP = 0$, the value of the pressure for which the flow is a maximum may be found. This pressure is defined as the critical pressure P_{crit} and is

$$P_{crit} = \frac{K}{E} \ldots \ldots \ldots \ldots \ldots [6]$$

Consequently, the maximum flow Q_{crit} can be defined in terms of the constants in Equation [5] and P_{crit}. Furthermore, if a relative pressure P_r and a relative flow Q_r are defined, a simple relation between these two variables can be derived

$$Q_r = \frac{2\sqrt{P_r}}{1 + P_r} \ldots \ldots \ldots \ldots \ldots [7]$$

where $Q_r = Q/Q_{crit}$ and $P_r = P/P_{crit} = PE/K$.

Fig. 3 shows a plot of Equation [7] together with some experimental results. It is to be observed that below P_r equal to unity, an increase in pressure results in an increase in flow. However, when P_r is greater than unity, an increase in pressure results in a decrease in flow. In this latter region of operation, the valve can be said to be behaving as a negative resistance, where the term resistance is understood to mean the slope of the pressure-flow characteristic curve

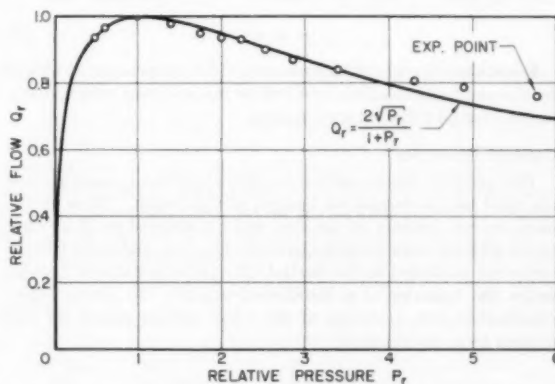$$\text{Resistance} = \frac{\partial P}{\partial Q}$$

Fig. 3   Comparison between theoretical and experimental valve characteristics

Significance of Negative Resistance

The influence of the valve negative resistance characteristic upon system stability can be visualized in the following manner: When a pressure wave traveling along the conduit reaches the valve, the latter (assuming that it has negligible mass or infinite natural frequency), if operating in the negative-resistance region, tends to reduce the flow instantaneously. This reduction in flow according to Euler's Equation brings about a pressure rise; hence an intensified reflected wave. Given enough time, the system will assume a steady oscillating condition. Fujii and Kiyama (4), using a valve having similar negative-resistance characteristics have solved the problem both analytically and graphically, and demonstrated how a small disturbance can grow when the supply pressure is above the critical pressure.

For the case when the valve mass and system friction cannot be neglected, a more complicated interaction results, and the valve may oscillate even though the steady-state value of P_r is less than unity. However, a closer look at the pressure fluctuations during the oscillations indicates that the dynamic value of P_r can go as high as two or three times the steady-state value where the resistance is negative. The significance of P_r of unity as a stability criterion is clouded in cases where complicated interactions take place or resonance occurs.

Dynamic-State Valve Analysis

Summing all the forces acting on the valve spool shown in Fig. 2, results in

$$m\ddot{X} + b\dot{X} + K[X - X_o] + EPX + \rho L\dot{Q} + F_c = 0 ..[8]$$

where the term $\rho L\dot{Q}$ is the damping force associated with the positive damping length as formulated by Lee (1).

With the subscript i denoting an initial steady-state condition and with the use of Equations [1] through [4], Equation [8] can now be written in a nondimensional form

$$\frac{\ddot{x}}{\omega_n{}^2} + \frac{2\zeta\dot{x}}{\omega_n} + x - 1 + r(px - 1) + \zeta_L q + f = 0 ... [9]$$

where

$$\omega_n = \sqrt{\left(\frac{K}{m}\right)} \quad x = \frac{X}{X_i} \quad \zeta = \frac{q}{2(Km)^{1/2}} \quad r = \frac{EP_i}{K} = \frac{P_i}{P_{crit}} = P_{ri}$$

$$p = \frac{P}{P_i} \quad q = \frac{Q}{Q_i} \quad \zeta_L = \frac{\rho LQ_i}{KX_i} \quad f = \frac{F_c}{KX_i}$$

Assuming that the orifice flow equation holds for the dynamic case, Equation [4] can be written in a nondimensional form

$$q = x\sqrt{p} ....[10]$$

Equations [9] and [10] describe the valve behavior as restricted by the assumptions made, and will be incorporated subsequently in the over-all valve-conduit system.

Conduit Dynamics

The spatially distributed nature of the fluid compressibility and the fluid inertia cannot be ignored in this study. Thus if the mean stream velocity of the fluid can be assumed small as compared with the velocity of sound in the medium, and if the friction terms are neglected in the partial differential equations that describe the behavior of a distributed conduit, the pressure-flow relationship just upstream of the valve can be solved for and written as in Equation [5]

$$2p_s\left(t - \frac{l}{c}\right) = p(t) + p\left(t - \frac{2l}{c}\right) + \phi q(t) - \phi q\left(t - \frac{2l}{c}\right)......[11]$$

where ϕ is the conduit nondimensional impedance which can be written as

$$\phi = \frac{\sqrt{(\rho\beta)}}{a} \frac{Q_i}{P_i}$$

For the derivation of Equation [11], the reader is referred to the many books in the field of transmission-line theory and water-hammer (6–9).

Since the supply pressure and the conduit pressure drop are assumed to be constant in this study, then

$$p_s\left(t - \frac{l}{c}\right)\Big|_{P_s = P_i} = 1 ....[12]$$

With this simplification, Equation [11] can be solved separately for the pressure and the flow at the valve end of the conduit for any time t

$$p(t) = 2 - p\left(t - 2\frac{l}{c}\right) - \phi q(t) + \phi q\left(t - 2\frac{l}{c}\right)..[13]$$

$$\phi q(t) = 2 - p(t) - p\left(t - 2\frac{l}{c}\right) + \phi q\left(t - 2\frac{l}{c}\right)..[14]$$

Figs. 4 (a and b) show block-diagram representations for Equations [13] and [14], respectively, in terms of time-delay components. For more elaborate treatment of this type of conduit representation, references (3) and (10) are recommended.
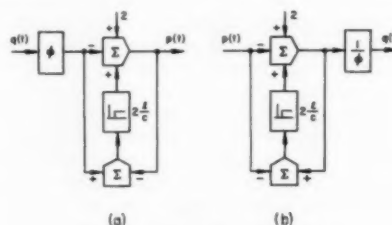
Fig. 4 Conduit block diagram
a (Equation [13]) b (Equation [14])

Valve Chamber Effect

The effect of the valve chamber, essentially that of a capacitance, is

$$q_l - q_o = \frac{\mathcal{U}_c P_i}{\beta_c Q_i}\dot{p} ....[15]$$

where

q_l = flow out of the conduit
q_o = flow through the orifice
$\mathcal{U}_c$ = valve-chamber volume between the spool lands
β_c = over-all bulk modulus of the fluid in the valve chamber

The presence of the valve chamber resulted in a system natural frequency substantially lower than the natural frequency of the conduit itself (11).

Valve, Conduit, Valve-Chamber System

Fig. 5 shows an analog-computer block diagram that represents the dynamical behavior of the valve, conduit, and valve-chamber system as characterized by the various equations given in the earlier part of this paper. In a study of the stability of this system (3), a signal f in the form of an impulse was applied to simulate an external disturbance. Observing the response of the valve position z, decides whether the system is stable or unstable under the specific operating conditions prescribed by a given set of parameters. A typical result of such a study together with some experimental points is shown in Fig. 6. The parameter r is the initial value of the relative pressure P_r, Equation [6]. It is to be pointed out that because of various interactions between the elements of the system, the graph does not reveal a definite change of trend as the value of r passes through unity, as would be anticipated from earlier discussions relating the negative re-

sistance idea. The parameter ω_n/ω_l is the ratio of the valve natural frequency to the quarter-wave-length frequency of the conduit. The latter is given by

$$\omega_l = \frac{\pi c}{2l}, \text{ rad/sec} \dots\dots\dots\dots\dots [16]$$

The graph in Fig. 6 indicates that there is clearly a state of resonance. The least stable point, however, does not occur at a value of ω_n/ω_l of unity because the quarter-wavelength frequency has been reduced by the presence of a valve chamber (11). From many observations the analog solutions agree well with experimental results both qualitatively and quantitatively. The largest discrepancies which are shown in Fig. 6 occur in the vicinity of the resonance point, where the disregard of the frictional effects in the conduit produces the largest errors.

Method of Valve Stabilization

The following methods are recommended for stabilizing a valve having a positive damping length:

1 *Increasing the Damping Length.* This seems to be the most powerful method to ensure valve stability before the valve is constructed. In 4-way valves, if negative damping lengths must be present, it is recommended that the total positive damping length should be at least three times the net negative damping length.

2 *Using a Composite Conduit.* An interesting method of stabilizing a valve is the use of a composite upstream conduit, a conduit made up of two or more sections having different cross-sectional areas and possibly different lengths. Both experimental and computer results have demonstrated the effectiveness of this technique. Fig. 7 shows dramatically how a composite conduit, if used properly, can stabilize a valve. In this specific case placing a short conduit (A) having a small cross-sectional area between the supply reservoir and another conduit (B) stabilized the valve.
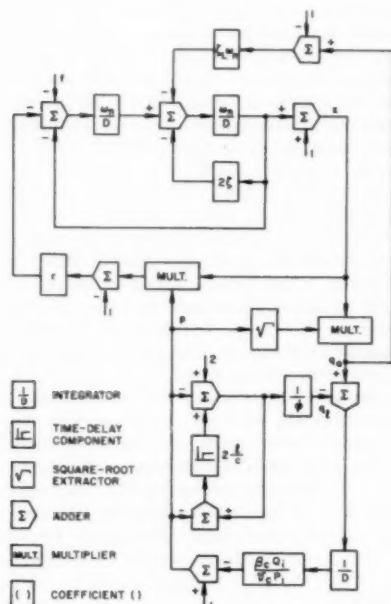
Fig. 5 Block diagram for a valve, conduit, valve-chamber system

Fig. 7 Qualitative effect of a composite conduit on valve stability

3 *The Connection of a Closed-End Conduit to Valve Chamber.* Often, a valve can be stabilized if a properly closed-end conduit, a Bourdon pressure gage, or an additional volume is connected to the valve chamber. The success of this method rests on changing the over-all natural frequency of the conduit system so that it will not be too close to the natural frequency of the valve itself.

4 *Decreasing the Value of r.* When all other parameters are held constant, decreasing the value of r makes the valve more stable

$$r = \frac{2C_q C_w w \cos\theta P_i}{K}$$

5 *Increasing the External Viscous Damping.* This method is to be used if all others fail.

Fig. 6 Comparison between experimental and analog values for r max for stability

Acknowledgment

References and Bibliography

1 "Contributions to Hydraulic Control—2, Transient-Flow Forces and Valve Instability," by S.-Y. Lee and J. F. Blackburn, TRANS. ASME, vol. 74, 1952, pp. 1013–1016.

2 "The Effect of Oil-Column Acoustic Resonance on Hydraulic Valve Squeal," by F. W. Ainsworth, TRANS. ASME, vol. 78, 1956, pp. 773–778.

3 "Effect of a Hydraulic Conduit With Distributed Parameters on Control Valve Stability," by F. D. Ezekiel, ScD thesis, Dept. of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Mass., June, 1955.

4 "Excitation of Vibration of Water Column by a Fluttering Valve (II)," by Sumiji Fujii and Gishiaki Kiyama, *Trans. of the Japan Society of Mechanical Engineering*, vol. 18, no. 73, 1952, p. 40.

5 Fluid Power Control Notes—Course 2.789, unpublished notes, Massachusetts Institute of Technology, Cambridge, Mass., 1956.

6 "Hydraulic Transients," by G. R. Rich, McGraw-Hill Book Company, Inc., New York, N. Y., 1955.

7 "Waterhammer Analysis," by J. Parmakian, Prentice-Hall, Inc., New York, N. Y., 1955.

8 "Analysis of Water Hammer by Characteristics," by C. A. M. Gray, *Trans. ASCE*, vol. 119, 1954, pp. 1176–1194.

9 "Theory of Water Hammer," by L. Allievi, English translation by Eugene E. Halmos, Engineering Societies, 1925.

10 "Computer Representations of Engineering Systems Involving Fluid Transients," by F. D. Ezekiel and H. M. Paynter, TRANS. ASME, vol. 79, 1957, pp. 1840–1850.

11 "Effect of an End Chamber on the Surge Frequency in a Hydraulic Conduit," by F. D. Ezekiel, ASME Paper No. 56—A-158(D).

Discussion

Frank W. Ainsworth.[4] The author has presented here a comprehensive treatment of an interesting phenomenon.

The problem of valve stability first came to my attention back in 1952 during the development of electrohydraulic servo-valves for high-performance autopilot servomechanisms. These valves incorporated the positive damping length feature (as discussed by the author) but, when tested on the bench, would often go violently unstable and wreck the plumbing. This condition was cured by adding viscous dampers to the valve spool, as discussed in the author's paper. However, the physics of the oscillation was still unexplained, and natural scientific curiosity demanded that it be investigated.

It was observed that the frequency of oscillation was well up into the acoustic range (500 cycles per second), which rough calculation showed to be close to the quarter wave length frequency of the attached oil column. This led to suspicion that the column of oil had a first order effect on the oscillation, which further analysis, as reported in the author's reference (2) proved true.

When we subsequently met to discuss our work, it was interesting to see the author's negative resistance and analog computer approach to the problem.

It will be interesting to see if in the future someone solves some serious valve stability problem with this new knowledge.

M. S. Feder.[5] This paper is a significant and fundamental con-

[4] Aero Division, Minneapolis-Honeywell Regulator Company, St. Petersburg, Fla.

[5] Research Engineer, Sperry Gyroscope Company, Great Neck, N. Y.

tribution to the field of hydraulic servo valves. The earlier work done by Lee on damping length has not explained many cases of valve instability observed in practice and for this reason the damping-length concept has neither been fully appreciated nor universally applied. This current work adds one more piece to the "valve-stability puzzle" and is important for this reason alone. Additionally, and perhaps of equal significance, this paper is noteworthy because it substantiates and strengthens Lee's original work and, in so doing, provides for a more rational approach to the valve-stability problem.

Although the current paper adds significantly to an understanding of valve stability, it is the writer's opinion that the two factors, damping length and conduit dynamics, are two of several factors which affect valve stability. Work done by the writer indicates that at least one additional factor affects stability of a single spool four-way valve. This is the variable pressure drop across the valve due to load dynamics. Regarding this point, tests on a servo valve driving a load showed that load inertia had a significant effect on valve stability. In effect, when a valve is used in practice (driving a load) the hydraulic spring rate is always variable and may cause valve instability. One solution to both the conduit problem and the "variable fluid spring" problem is to swamp out fluid forces with a large stiff torque motor. Another is to utilize a two-stage servo valve. Both solutions, in effect, increase available driving force and servo-valve stiffness. With regard to valve stiffness and its effect on stability, it is interesting to note, that the present paper may explain why fluid force "compensated" valves are notoriously difficult to stabilize. These valves use soft torque motors. Static fluid forces in theory can be made zero by proper contouring of the valve spool. In actual practice cancellation of fluid forces is not complete particularly near the valve null. In this region (as shown in Fig. 6) r will be large and the valve will tend toward instability.

With regard to the current paper, the analyses show that conduit dynamics have a significant effect on valve stability. The results presented in Fig. 6 are perhaps the most significant in the report. However, these results, although not completely general, would be more useful if either data for a range of valves were shown or a discussion concerning the applicability of Fig. 6 were presented. Without this additional information a separate analog computer study may be required for every valve design.

Author's Closure

The author would like to thank Messrs. Ainsworth and Feder for their interesting comments and helpful suggestions. Mr. Feder is quite correct in pointing out that the damping length and conduit dynamics are only two of several factors which effect valve stability in a general system.

This paper focuses attention on a very simple situation in which the interaction between the valve and the conduit are brought out. In a more complete system, load effects should, of course, be included and would undoubtedly affect the system behavior.

Although the results shown in Fig. 6 are only applicable to a valve having a single metering edge, yet it may be deduced that in a more complicated situation it is advisable to avoid close proximity between the valve natural frequency and that of the conduits. In evaluating the latter, reference (11) may be of some help.

In closing, the author hopes that further work may shed more light on the rather elusive problem of valve stability.

Work Capacities of Energy Storage Systems on Basis of Unit Weight and Unit Volume

By L. V. KLINE,[1] S. M. MARCO,[2] AND W. L. STARKEY[3]

In view of current and anticipated interest in conserving weight and space in engineering applications, a study is made of work-per-unit-weight and work-per-unit-volume moduli for energy storage systems. Gasoline engines, solid propellants, electrochemical cells, flywheels, compressed gases, compressed liquids, and stressed solids are evaluated. General analytical expressions for the work moduli are derived. For each such general expression, significant optimizing parameters are selected and methods of optimizing the moduli are presented. Information is included pertinent to the problem of selection of energy storage systems for specific applications. Comparisons are made of the practical maximum values of weight and volume moduli for the considered systems. Limitations of the systems, including maximum allowable stress, are discussed.

NOMENCLATURE

The following nomenclature is used in the paper:

b = width of flywheel rim measured along axis which is parallel to axis of rotation or width of spring

c_v = specific heat at constant volume

C = dimensionless thickness function b/x_2 or constant defined in text

d = wire diameter

d_i = inside diameter

d_o = outside diameter

D = mean coil diameter

E = work capacity

E_m = modulus of elasticity

g = acceleration conversion factor

G = modulus of elasticity in shear

h = height of Belleville spring

I = moment of inertia

L = ln x_1/x_2 or length

L_c = length of cylindrical portion of container

L_f = free length of spring

M = mass

n = number of arms in rim-arm flywheel or number of coils in coil spring

N_1 = number of groups of springs which are stacked in series in a stack of Belleville springs

N_2 = number of springs which are stacked in parallel in any one group in a stack of Belleville springs

p_f = pressure after expansion has occurred

p_0 = pressure before expansion has occurred

r = radius

r_a = radius of arbor on which spiral-wound spring is wound

r_c = radius of casing which houses clock-type spring

r_m = mean radius of rim

r_1 = inside radius of flywheel rim or of cylinder

r_2 = outside radius of flywheel rim, of cylinder, or of spring

R = universal gas constant

S_m = maximum normal stress

S_s = maximum shearing stress

t = thickness of rim measured along axis perpendicular to axis of rotation, thickness of cylindrical-shaped disk or thickness of spring

T_1 = temperature of gas before expansion has occurred

V = volume of space occupied

W = weight

W_m = weight of working medium

x = thickness of uniform-stress disk measured along an axis parallel to axis of rotation

x_1 = thickness of uniform-stress disk at zero radius

x_2 = thickness of uniform-stress disk at radius $r = r_1$

β = radius ratio r_1/r_2

γ = radius ratio r_2/r_1 or ratio of specific heats

δ = thickness ratio x_1/x_2

μ = Poisson's ratio

ρ = density

ρ_c = density of container material

ρ_m = density of working medium

ω = angular velocity

INTRODUCTION

For most engineering applications, increasing emphasis is being placed on conservation of weight and volume. This emphasis is the logical result of increased operating speeds, the necessity for economy in competitive markets and, as in the case of aircraft, is frequently a fundamental requirement to insure successful operation. It is felt, therefore, that information concerning the weight and volume moduli of energy storage systems is valuable in broadening the base of knowledge in the field of machine design. Weight modulus is defined to be the work capacity of an energy system per unit weight of the system. Volume modulus is defined to be the work capacity per unit of space occupied by the system. This study includes general analytical expressions for these moduli, serves as a guide in selecting components of energy storage systems, and supplies information on minimizing the weight or space occupied by the components. The energy storage systems which are studied are those most commonly used in engineering.

ANALYSIS

Consider that it is desired to evaluate the weight modulus for a flywheel. If it is assumed that the work capacity is equal to the kinetic energy of the flywheel, the work capacity can be expressed as a function of the angular velocity, the dimensions, and the density of the flywheel material. Thus

$$E = f(\omega, \rho, D)$$

where ω, ρ, and D indicate angular velocity, density, and a dimension function, respectively. Since the angular velocity can be expressed as a function of the dimensions and the maximum stress in the flywheel, it is possible to rewrite the work capacity expression as

$$E = \varphi(D_1, S_m, \rho)$$

where D_1 is some dimension function different from D. The restriction on the stress S_m is that it remain less than or equal to the yield-point stress of the working material. Dividing the work-capacity expression by the weight of the flywheel gives the weight modulus as a function of dimensions, density, and maximum material stress. In a similar fashion, dividing the work-capacity expression by the volume of space occupied by the flywheel gives the volume modulus. This type of analysis is carried out for several of the systems commonly used in engineering. The resulting analytical expressions are given in the Appendix. In each case the moduli are expressed in terms of material stresses. In the case of compressed fluids, the material stress is the stress in the container, since it is this stress which controls the size and weight of the container and therefore of the energy storage system.

The general equations can be examined to determine what parameters will maximize the moduli. A brief summary outlining the optimum parameters follows each set of equations in the Appendix.

By using the combination of parameters which maximizes the moduli for each system, assigning reasonable numerical values to these parameters, and using a common value for the maximum stress in each system, it is possible to obtain a comparison between the maximum values of the moduli. Such a comparison is included as Tables 1 and 2. The normal stress which is chosen for the compilation of the tables is $S_m = 100,000$ psi, for all systems. The maximum shear stress is considered to be one half this value. Several common energy storage systems other than those of the Appendix are included for comparison purposes.

General Observations

The gasoline engine is one of the best work-producing devices from the standpoint of weight and volume modulus. The weight modulus for a gasoline engine is about 4.5 times that for the next best work-producing device considered. The volume modulus is about 6.5 times greater than that of the silver-zinc electrochemical cell. These results are due to the energy being stored in a highly concentrated form in the fuel. This fact is emphasized by including in Tables 1 and 2 the heating value of the gasoline.

If electrical energy is desired as the output of the system, and it is unnecessary to extract energy more rapidly than it can be extracted from a silver-zinc electrochemical cell, this cell is recommended as an energy storage device.

If it is desired to extract energy from the system more rapidly than it can be extracted from an electrochemical cell and the use of a gasoline engine is not practical, then a flywheel system provides the best moduli. The weight and volume moduli for flywheels can be increased over that obtained for a cylindrical flywheel by shaping the flywheel as a uniform-stress disk with rim attached. This increase however may not be sufficient to justify the added manufacturing costs.

If it is desired to store energy for longer periods of time than is practical in a flywheel, and it is necessary to extract energy more rapidly than it can be extracted from an electrochemical cell, and the use of a gasoline engine is not practical, then a compressed-gas system is best. Spherical containers for compressed-gas systems are superior to cylindrical containers from the standpoint of weight and volume moduli.

Springs are inferior to the gasoline engine by a factor of 13,000

TABLE 1   OPTIMUM WEIGHT-MODULUS COMPARISONS[a]

System	Ft lb/lb
Gasoline (fuel alone)[b]	14500000
Solid propellant[c]	1050000
Gasoline engine[d]	660000
Ag-Zn electrochemical cell[e]	146000
Lead-acid electrochemical cell[e]	38000
Edison electrochemical cell[e]	35000
Uniform-stress disk with rim (Fig. 1, $C = 1$, $\delta = 4$)	26700
Compressed gas (spherical container)[f]	22600
Compressed gas (cylindrical container)[f]	18400
Cylindrical flywheel (Equation [2], $t/r = 0.2$)	18000
Rim-arm flywheel[g]	2450
Compressed liquid (ether)[h]	300
Compressed solid (steel)[i]	49
Compressed solid (torsion spring) (Equation [5], $d_i/d_o = 1$)	29
Compressed solid (spiral-wound spring) (Equation [6])	16
Compressed solid (coil spring) (Equation [4])	15
Compressed solid (Belleville spring)[j]	11

[a] Common values of constants which were used in the evaluations of Tables 1 and 2 are: $\rho = 7.31 \times 10^{-4}$ lb sec^2/in^4, $G = 11.5 \times 10^6$ psi, $\mu = 0.288$, $E = 30.0 \times 10^6$ psi, $S_o = 50,000$ psi, $S_m = 100,000$ psi, and $\rho g = 0.283$ pci.
[b] Evaluated from typical heating values.
[c] Evaluated from typical heats of explosion.
[d] Evaluated by assuming volume occupied by fuel is equal to volume occupied by engine, density of fuel is 50 pcf, equivalent density of engine is 250 pcf, and specific fuel consumption is 0.5 lb/bhphr.
[e] Evaluated from representative manufacturer's information.
[f] From Fig. 5 where $p_o = 6000$ psi and $p_f = 14.7$ psi.
[g] From Equations [3] where $t/r_m = 0.3$, $n = 8$, and $S_m = 25,000$ psi. The allowable maximum stress in the rim-arm-type flywheel has been chosen lower than in the other types of flywheels owing to the greater stress concentrations.
[h] From Equations [10] and combination gaphical-analytical solution where $p_o = 2000$ atm, $p_f = 1$ atm, and $Lc/r_1 = 10$.
[i] From basic energy relations for compressing a solid block of steel with uniaxial force.
[j] From Fig. 3 where $t/r_2 = 0.1$ and $\beta = 0.8$.

TABLE 2   OPTIMUM VOLUME-MODULUS COMPARISONS[a]

System	$\dfrac{\text{Ft lb}}{\text{ft}^3} \times 10^{-2}$
Gasoline (fuel alone)[b]	682000
Solid propellant[c]	105000
Gasoline engine (see note (d), Table 1)	99000
Ag-Zn electrochemical cell[d]	14800
Uniform-stress disk with rim (Fig. 2, $C = 1.5$, $b = x_1$)	8780
Cylindrical flywheel (Equation [2], $t/r = 0.2$)	8750
Lead-acid electrochemical cell[d]	5000
Edison electrochemical cell[d]	4800
Compressed gas (spherical container)[e]	2500
Compressed gas (cylindrical container)[e]	2500
Rim-arm flywheel[f]	1070
Electrostatic capacitor[d]	200
Compressed liquid (ether)[g]	90
Compressed solid (steel)[h]	24
Compressed solid (torsion spring) (Equation [5], $d_i/d_o = 1$)	8
Compressed solid (coil spring) (Equation [4], $D/d = 2$)	5
Compressed solid (spiral-wound spring) (Equation [6], $r_o/r_c = 0.1$)	4
Compressed solid (Belleville spring)[i]	2

[a] See note (a), Table 1.
[b] Evaluated from typical heating values and densities.
[c] Evaluated from typical heats of explosions and densities.
[d] Evaluated from representative manufacturer's information.
[e] From Fig. 6 where $p_o = 10,000$ psi and $p_f = 14.7$.
[f] From Equations [3] where $t/r_m = 0.3$ and $n = 8$; see note (g), Table 1 for explanation concerning S_m.
[g] From Equations [10] and combination gaphical-analytical solution where $Lc/r_1 = 10$, $p_o = 2000$ atm, $p_f = 1$ atm.
[h] See note (i), Table 1.
[i] From Equations [4] where $t/r_2 = 0.1$ and $\beta = 0.625$.

on basis of weight modulus and by a factor of 4000 on basis of volume modulus. Springs are inferior to a compressed-gas system by a factor of 450 based on the weight modulus and by a factor of 100 based on the volume modulus.

This study emphasizes that the relative merits of space con-

servation and weight conservation must be evaluated for any design. Thus, for example, for a uniform-stress disk with rim attached, the combination of dimensions which produces maximum energy per unit weight is not the combination of dimensions which produces the maximum energy per unit volume. For a torsion-bar spring the dimensions which produce the maximum weight modulus are the dimensions which produce minimum volume modulus.

The limitations of the use of weight and volume moduli are recognized. These moduli are only two of the many factors which need to be considered in any design. Other factors are:

1   Economy.
2   Reliability.
3   Ease of manufacturing required components.
4   Rate at which energy must be available.
5   Time for which the energy must be stored.
6   Energy losses associated with system.
7   Amount of maintenance required.
8   Auxiliary equipment necessary.
9   Operation ability under extremes of temperature, dust, and humidity.
10   Type of energy needed as output from the system.

Thus, although a spring system is the worst of those considered from the standpoint of weight and volume moduli, a spring system may be best from the standpoint of time for which the energy can be stored. If electrochemical cells are being considered for an application, it must be realized that, although they have excellent weight and volume moduli, they may require a considerable amount of auxiliary equipment to supply electrical energy with the necessary characteristics. A system with a lower weight modulus and fewer auxiliary components might be more desirable.

Obviously, losses in the system are important. Although the compressed-liquid system appears to be superior to the compressed-solid systems on the basis of weight and volume moduli, the valve losses and the losses resulting from temperature changes make its use questionable in a practical system. It is realized, therefore, that this study does not supply the final answer as to what energy storage system is most suitable for a given application, but it is hoped that it will supply basic information which will aid the engineer in making intelligent decisions.

Appendix

GENERAL ANALYTICAL EXPRESSIONS AND OPTIMUM PARAMETERS

The final equations, the primary assumptions used in deriving them, and a brief discussion of the optimum parameters are included in this section. Space limitations prohibit inclusion of the detailed analyses.

Uniform-Stress Disk

Assumptions. Bending stresses in the rim are neglected. Plane-stress conditions exist in the rim and disk. Volume of space occupied is $V = \pi b r_2^2$ if $b > x_1$, and $V = \pi x_1 r_2^2$ if $b < x_1$. Work capacity equals total kinetic energy

$$\frac{E}{W} = \frac{S_m\left[\dfrac{\delta - L - 1}{L} + \dfrac{CL(\gamma^4 - 1)}{2}\right]}{g\rho\left[\dfrac{\delta - 1}{L} + C(\gamma^2 - 1)\right]}$$

$$\frac{E}{V} = S_m\left[\frac{\delta - L - 1}{LC\gamma^2} + \frac{L(\gamma^4 - 1)}{2\gamma^2}\right] \quad \text{for} \quad b > x_1$$

$$\frac{E}{V} = S_m\left[\frac{\delta - L - 1}{\delta L\gamma^2} + \frac{bL(\gamma^4 - 1)}{2\gamma^2 x_1}\right] \quad \text{for} \quad b < x_1$$

The radius ratio γ is a dependent variable given by

$$\gamma = \left\{\frac{\varphi}{2} + \left[\left(\frac{\varphi}{2}\right)^2 - \theta\right]^{1/2}\right\}^{1/2} \qquad \dots[1]$$

where

$$\theta = \frac{\mu - 1}{3 + \mu}\left[\frac{2\left(\dfrac{1}{C} - 1\right)}{L} + 1\right]$$

and

$$\varphi = \frac{2}{3 + \mu}\left[1 + \mu + \frac{\mu + 1}{LC} + \frac{1 - \mu}{L}\right]$$

These conditions are necessary to insure uniform stress in the disk portion of the flywheel.

Optimum Parameters. (Figs. 1 and 2) To maximize weight modulus, C should be unity, and γ should be as close to unity as the limiting speed of the flywheel permits. To maximize volume modulus, b should equal x_1. For any value of C, there is a corresponding value of γ which will maximize volume modulus. Lines of constant δ are not drawn on Fig. 2. However, for any combination of C and γ, the associated value of δ can be read from Fig. 1.

Rotating Plane Cylinder

Assumptions. Work capacity is equal to total kinetic energy

$$\frac{E}{W} = \frac{2S_m}{\rho g\left[3 + \mu + \dfrac{\mu(1 + \mu)(t/r)^2}{3(1 - \mu)}\right]}$$

$$\frac{E}{V} = \frac{2S_m}{3 + \mu + \dfrac{\mu(1 + \mu)(t/r)^2}{3(1 - \mu)}} \qquad \dots\dots[2]$$

Optimum Parameters. To maximize the moduli, t/r should be as small as possible. The limiting value of weight modulus as t/r approaches 0 is

$$\frac{E}{W} = \frac{S_m}{\rho g(1.644)}$$

for the case where $\mu = 0.288$. The limiting value of volume modulus as t/r approaches 0 is

$$\frac{E}{V} = \frac{S_m}{1.644}$$

Rim-Arm Flywheel

Assumptions. All of energy is stored in the rim. Stretch of arms is three fourths of that necessary for free expansion of the rim. Stress in the rim due to centrifugal force is found by assuming the portion of the rim between the arms to be a straight beam
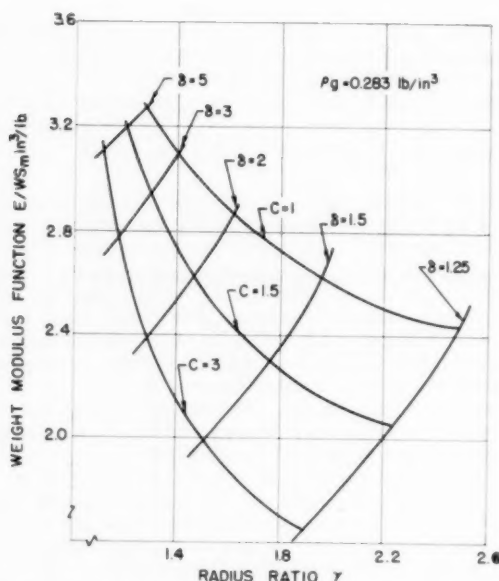
FIG. 1   WEIGHT MODULUS FOR UNIFORM-STRESS DISK
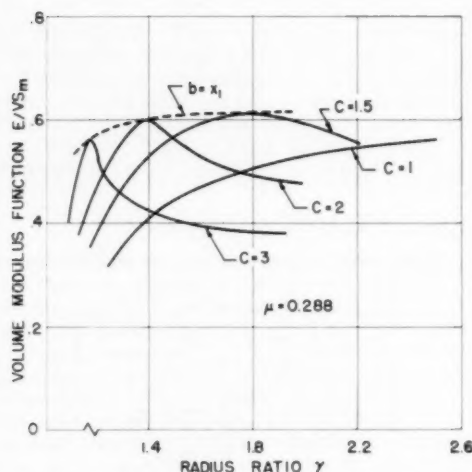
FIG. 2   VOLUME MODULUS FOR UNIFORM-STRESS DISK

uniformly loaded with the centrifugal force. Mass of the flywheel is 1.5 times the mass of the rim. Volume of space occupied is $V = \pi b r_m^2$. Work capacity is equal to total kinetic energy

$$\left.\begin{array}{l} \dfrac{E}{W} = \dfrac{S_m}{3\rho g \left(0.75 + \dfrac{\pi^2 r_m}{2n^2 t} \right)} \\[4ex] \dfrac{E}{V} = \dfrac{S_m}{\left(\dfrac{r_m}{t} \right) \left(0.75 + \dfrac{\pi^2 r_m}{2n^2 t} \right)} \end{array}\right\} \quad \dots [3]$$

Optimum Parameters. To maximize the moduli, t/r_m should be as large as the assumptions permit. Similarly, the number of arms should be the largest that the assumptions permit. This

results because the larger number of arms permits a higher speed to be attained before the rim becomes overstressed.

Coil Spring

Assumptions. Work capacity is equal to elastic resilience. Length of spring wire is $l = \pi D n$. Maximum stress in spring occurs when spring is compressed solid. Volume of space occupied is

$$V = \pi (D + d)^2 L_f / 4$$

$$\left.\begin{array}{l} \dfrac{E}{W} = \dfrac{S_s^2}{4g\rho G} \\[4ex] \dfrac{E}{V} = \dfrac{S_s^2 \pi}{4 \left(\dfrac{D}{d} \right) \left(\dfrac{D}{d} + 1 \right)^2 \left[G \left(\dfrac{d}{D} \right)^2 + S_s \pi \right]} \end{array}\right\} \dots [4]$$

Optimum Parameters. The weight modulus is independent of the dimensions of the spring. The volume modulus increases as D/d decreases, as would be expected.

Hollow Torsion-Bar Spring

Assumption. Volume of space occupied is $V = L \pi d_o^2 / 4$. Work capacity is equal to elastic resilience

$$\left.\begin{array}{l} \dfrac{E}{W} = \dfrac{S_s^2 \left[1 + \left(\dfrac{d_i}{d_o} \right)^2 \right]}{4\rho g G} \\[4ex] \dfrac{E}{V} = \dfrac{S_s^2 \left[1 - \left(\dfrac{d_i}{d_o} \right)^4 \right]}{4g} \end{array}\right\} \quad \dots \dots \dots [5]$$

Optimum Parameters. The value of d_i/d_o which maximizes the weight modulus minimizes the volume modulus. The weight modulus approaches a maximum limit of

$$\frac{E}{W} = \frac{S_s^2}{2g\rho G}$$

as d_i/d_o approaches its physical limit of unity. The volume modulus approaches a maximum limit of

$$\frac{E}{V} = \frac{S_s^2}{4G}$$

as d_i/d_o approaches zero. This is the solid-bar case.

Flat Spiral-Wound Clock-Type Spring

Assumptions. Bending moment is uniform along length of spring. Spring is of rectangular cross section. Weight of spring casing can be neglected. Volume of space occupied is $V = \pi b r_c^2$. The spring occupies one half the available space in the drum. Work capacity is equal to elastic resilience

$$\left.\begin{array}{l} \dfrac{E}{W} = \dfrac{S_m^2}{6g\rho E_m} \\[4ex] \dfrac{E}{V} = \dfrac{S_m^2 \left[1 - \left(\dfrac{r_a}{r_c} \right)^2 \right]}{12 E_m} \end{array}\right\} \quad \dots \dots \dots [6]$$

Optimum Parameters. The weight modulus is independent of the dimensions of the spring. To maximize volume modulus, r_a/r_c should be minimum.

Belleville Spring

Assumptions. When N_2 springs are stacked in parallel, the load capacity is N_2 times the load capacity of a single spring for a given deflection. When N_1 groups of springs are stacked in series, the deflection is N_1 times the deflection of a single spring for any given load. The load deflection and stress-deflection relations are given by Almen and Laszlo.[4] Maximum stress in spring occurs when a single spring undergoes a deflection equal to its height. The volume of space occupied by a single vertical stack of springs is $V = \pi r_2^2 N_1[h + N_2 t]$. Work capacity is equal to elastic resilience

$$\left.\begin{array}{l} \dfrac{E}{W} = \dfrac{E_m(t/r_2)^4(h/t)^2[(h/t)^2 + 4]}{8g(1 - \mu^2)\pi\rho C(1 - \beta^2)} \\[3mm] \dfrac{E}{V} = \dfrac{E_m(t/r_2)^4(h/t)^2[(h/t)^2 + 4]}{8(1 - \mu^2)C\pi[(h/tN_2) + 1]} \end{array}\right\} \quad \ldots\ldots\ldots [7]$$

where

$$\frac{h}{t} = \frac{C_2}{C_1}\left\{\left[1 + \frac{2S_m(1 - \mu^2)C\ C_1(r_2/t)^2}{E_m C_2^2}\right]^{1/2} - 1\right\}$$

and

$$C = \frac{6\left(\dfrac{\beta - 1}{\beta}\right)^2}{\pi \ln \beta}$$

$$C_1 = \frac{6\left(\dfrac{\beta - 1}{\ln \beta} - 1\right)}{\pi \ln \beta}$$

$$C_2 = \frac{6\left(\dfrac{\beta - 1}{2}\right)}{\pi \ln \beta}$$

Optimum Parameters (Figs. 3 and 4). Since the foregoing equation for weight modulus is independent of N_1 and N_2, Fig. 3 is valid for a stack of Belleville springs as well as for a single spring. The effect of increasing β and t/r_2 can be seen from this figure. Since the volume-modulus equation is independent of N_1, Fig. 4

4 "The Uniform-Section Disk Spring," by J. O. Almen and A. Laszlo, Trans. ASME, vol. 58, 1936, pp. 305–314.

is valid for a series stack of springs as well as for a single spring. This figure reveals that for any given t/r_2 there is a value of β which maximizes the volume modulus.

Compressed Gas (Cylindrical Container)

Assumptions. Work-extraction process is isentropic. Piston-cylinder device is used for extracting work. Gas obeys perfect gas law. Weight of ends of container can be neglected

$$\left.\begin{array}{l} \dfrac{E}{W} = \dfrac{\dfrac{c_v}{R}\ [1 - (p_f/p_0)^{(\gamma - 1)/\gamma}]}{\dfrac{1}{RT_1} + \dfrac{g\rho_e}{S_e - p_0}} \\[5mm] \dfrac{E}{V} = \dfrac{p_0 c_v}{R}\ [1 - (p_f/p_0)^{(\gamma - 1)/\gamma}]\left[1 - \dfrac{p_0}{S_e}\right] \end{array}\right\} \ldots [8]$$

Optimum Parameters (Figs. 5 and 6). Air has been assumed as the working medium in these curves. If air is expanded from the storage pressure to atmospheric pressure, there is some value of storage pressure which optimizes the weight modulus. However, if the air is expanded through a given pressure ratio, the weight modulus is nearly constant over a wide pressure range. The effects of the parameters on volume modulus can be seen from Fig. 6. The difference in the moduli for the spherical container and the cylindrical container is too small to be read on this curve.

Compressed Gas (Spherical Container)

Assumptions. Work-extraction process is isentropic. Piston-cylinder device is used for extracting work. Gas obeys perfect gas law

$$\left.\begin{array}{l} \dfrac{E}{W} = \dfrac{\dfrac{c_v}{R}\ [1 - (p_f/p_0)^{(\gamma - 1)/\gamma}]}{\dfrac{1}{RT_1} + \dfrac{3\rho_e g}{4S_e - 3p_0}} \\[5mm] \dfrac{E}{V} = \dfrac{p_0 c_v}{R}\ [1 - (p_f/p_0)^{(\gamma - 1)/\gamma}]\left[1 - \dfrac{3p_0}{4S_e}\right] \end{array}\right\} \ldots [9]$$

Optimum Parameters. Refer to Figs. 5 and 6 and the discussion pertaining to a cylindrical container.

FIG. 3 WEIGHT MODULUS FOR BELLEVILLE SPRING
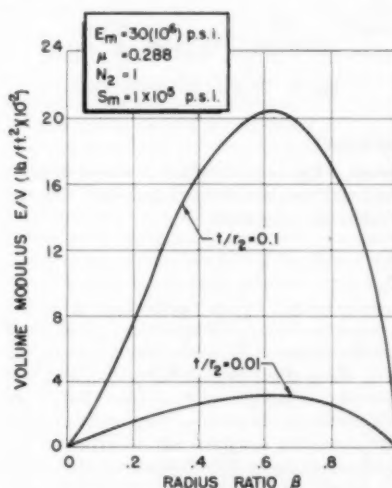
FIG. 4 VOLUME MODULUS FOR BELLEVILLE SPRING
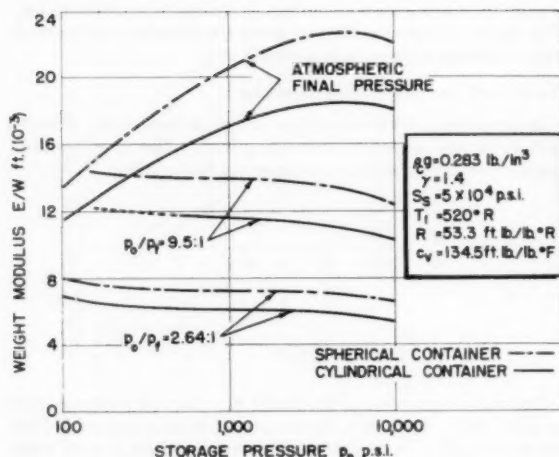
FIG. 5 WEIGHT MODULUS FOR AIR
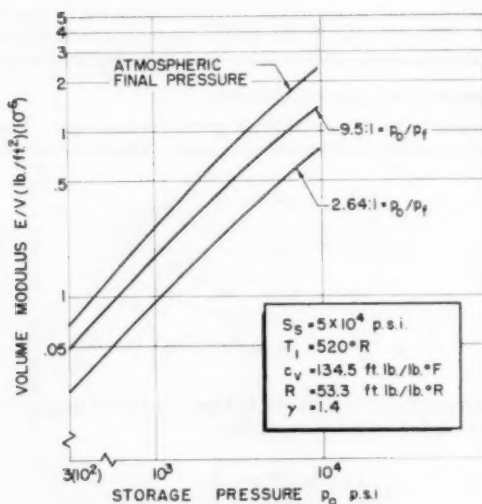
FIG. 6 VOLUME MODULUS FOR AIR

Compressed Liquid

Assumptions. Compressed liquid is stored in piston-cylinder arrangement with hemispherical-shaped end. Energy stored in container material is negligible

$$\frac{E}{W} = \frac{E}{W_m}$$

$$\left\{ \frac{\dfrac{L_c}{r_1} + \dfrac{2}{3}}{\dfrac{L_c}{r_1} + \dfrac{2}{3} + \dfrac{\rho_s}{\rho_m}\left[\dfrac{L_c(1-\alpha^2)}{\alpha^2 r_1} + \dfrac{2(1-\alpha^3)}{3\alpha^3}\right]} \right\} \quad ..[10]$$

$$\frac{E}{V} = \frac{E}{V_m} \left\{ \frac{\dfrac{L_c}{r_1} + \dfrac{2}{3}}{\dfrac{L_c}{r_1} + \dfrac{2}{3} + \dfrac{L_c(1-\alpha^2)}{\alpha^2 r_1} + \dfrac{2(1-\alpha^3)}{3\alpha^3}} \right\}$$

where

$$\alpha^2 = 1 - \frac{p_0}{S_s}$$

Optimum Parameters. Before the optimum parameters can be determined, information must be obtained concerning E/W_m and E/V_m. For this study ether has been chosen as the working medium. Data for the pressure-relative volume ratio has been obtained from Bridgman[5] for an isothermal process. Utilizing these data and a combination graphical-analytical solution, E/W_m and E/V_m can be calculated. For any given value of final pressure there is some value of storage pressure which maximizes the weight and volume modulus.

Discussion

R. M. ROTTY.[6] This paper presents an interesting approach to the evaluation of energy storage systems. The unique parameters of weight-modulus and volume-modulus provide additional information which too often is not considered.

However, in any design the weight and volume moduli are only two of many factors which must be considered. This point was appropriately made by the authors, and the other factors listed all appear to be appropriate. It seems that the type of energy needed should far outweigh most of the others in importance. If mechanical work is assumed to be the most desirable form of energy output from the storage system, then the tremendous amounts of energy that are stored in gasoline, in solid propellants, and in the nucleus of fissionable material cannot be compared with other energy-storage systems mentioned here, except on an entirely different basis.

Perhaps a tabulation of the approximate fraction of the energy stored which is "available" to do mechanical work along with the weight and volume moduli would be appropriate. Nearly all the other forms of energy for which storage systems are suggested can be quickly and efficiently transformed into heat energy, but, of course, the second law of thermodynamics points out that the reverse is not true.

The title of this paper suggests that the storage of the energy is to be reversible. That is, if a certain quantity of mechanical energy is being delivered in excess of that which is actually needed, let us attempt to store this excess energy for future use. In order for a process of this type to be feasible, it is necessary for the storage process to be approximately reversible. Certainly, the compression of any fluid (liquid or gas), the compression of a spring, the addition of energy to a flywheel, and many others meet this requirement. Any fuel and any fuel-burning engine are in an entirely different category. These are examples of irreversible energy storage. The energy in the fuel was put there by Nature (and can be duplicated by man only through tremendous expenditures and waste of other forms of energy) and the release of this energy, regardless of its form, is irreversible and hence these systems cannot be used to "store" excess energy.

Certainly these facts were evident to the authors, and probably gasoline, solid propellants, and gasoline engines are listed only to give a comparison of the energy quantities in a certain mass or in a certain volume. As a comparison, it appears more in order that these be presented in a separate table along with a few other typical fuels and other types of prime movers. Such a tabulation perhaps could be made to distinguish between the forms of energy release; i.e., heat and mechanical energy, at the same time.

[5] "The Physics of High Pressure," by P. W. Bridgman, Bell and Sons, Ltd., London, England, 1931, p. 128.
[6] Associate Professor of Mechanical Engineering, Michigan State University, East Lansing, Mich. Mem. ASME.

Resistance to Rolling and Sliding[1]

By A. C. DUNK[2] AND A. S. HALL, JR.,[3] LAFAYETTE, IND.

In general a single "coefficient of friction" is inadequate to define the frictional resistance between contacting bodies. It is necessary to know both the inclination of the resultant force line and the offset of this line from the theoretical point of contact. This paper presents results of experiments, which were made to determine the effects of velocity on the inclination, and the offset of the line of action of the resultant force between two disks in peripheral contact.

The results indicate that the sliding velocity is of primary importance in determining the inclination of the line of action of the resultant force. It is the sum of the sweep velocities that is of primary importance in determining the offset of the resultant force line. The nature of the relationship between the angle of inclination and the offset of the line of action of the resultant force, and the proximity of the contact conditions to those existing for pure rolling, was investigated but no conclusive results were obtained.

NOMENCLATURE

The following nomenclature is used in the paper:

W = load applied along line of centers, Fig. 1
P = resultant load between disks per unit length of line contact
Z = viscosity of lubricant
V_s = sliding velocity, equal to algebraic difference of sweep velocities
V_Σ = sum of sweep velocities
n = ratio of numerically larger to smaller sweep velocities
S = rolling contact proximity factor equal to $[(1-n)/n]^2$
R = relative radius of curvature of contacting profiles
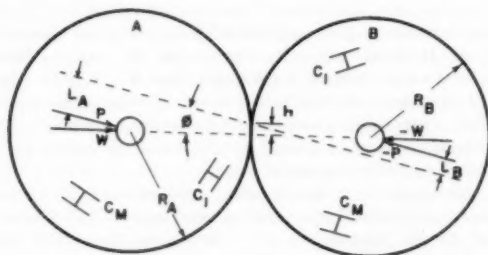ϕ = angle of inclination of line of action of resultant force between disks, with line of centers

FIG. 1 NOTATION

h = offset of line of action of resultant force, measured perpendicular to line of centers, from theoretical point of contact
h_r = offset h when disks are in rolling contact
C_M, C_I, C_W, C_B = respectively, motor, idling, windage, and bearing couples applied to disk

INTRODUCTION

The frictional resistance in some important special situations, notably journal bearings with thick-film lubrication, has been investigated both experimentally and theoretically with considerable success. Much remains to be learned about the more general situation of contact between bodies of any shape in relative motion. There is a large number of variables in any contact situation and it is difficult to separate the effects of each, in an attempt to obtain a correlation among the important variables.

Merritt[4] has made a study of the friction between a pair of disks mounted on parallel shafts and in peripheral contact, in an attempt to determine the nature of the frictional resistance between a worm and a wheel in a speed reducer. The results are plotted as a coefficient of friction versus the sliding velocity for different ratios of the linear velocities of the disks, and for different lubricants. Merritt defines his coefficient of friction in terms of the energy loss, the applied load, and the sliding velocity between the disks. This definition leads to the rather disconcerting result that the coefficient of friction becomes infinite as the sliding velocity approaches zero. In addition, the results of Merritt's tests do not yield exact information regarding the location of the resultant force between the disks. No distinction or separation of possible rolling and sliding resistances is possible.

PROBLEM ANALYSIS

For two bodies executing plane relative motion, the center of relative rotation (instant center of relative velocity) may be anywhere along the normal to the contacting curves passing through the point of contact. Motion of this type, for example, is present between mating spur gears, between cam and follower, and between elements of journal and antifriction bearings. If the surfaces are continuous and no fluid film exists in the contact region, then point or line contact will be approached for small loads. The line of action of the resultant force between the surfaces for the assumed case of point or line contact may be inclined to the line joining the centers of curvature but it must pass through the point of contact. If the load is increased sufficiently to cause deformation or if a fluid film is added to the contact region, the line of action of the resultant force may now be displaced as well as rotated with respect to the line joining the centers of curvature. In general, then, two quantities must be known in order to specify the location of the line of action of the resultant force between the surfaces relative to the common normal; viz., its linear displacement and its angular displacement.

From the foregoing analysis of the contact between curved surfaces having plane relative motion, it is evident that the effects of the variables on the frictional resistance could be studied in terms of the line of action of the resultant force between the surfaces. This paper is a report on research carried out to determine the effects of velocity on the frictional resistance between surfaces in terms of the resultant force between the surfaces.

[4] "Worm Gear Performance," by H. E. Merritt, Proceedings of The Institution of Mechanical Engineers. vol. 129, 1935, pp. 127–194.
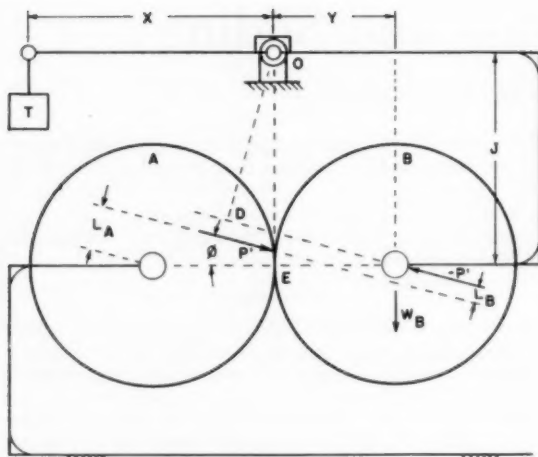
FIG. 2 SKELETON DIAGRAM OF APPARATUS SHOWING RESULTANT
FORCE P BETWEEN DISKS

FIG. 3 FREE-BODY DIAGRAM OF DISK B

THE INVESTIGATION

The apparatus consisted of a pair of disks separately mounted on low-friction bearings and independently motor driven through spiral springs which were calibrated, and were used to measure the motor torque supplied to each disk. Rotating protractors and indexes, read with the aid of synchronized stroboscopes, were used to measure the angles of twist of the spiral springs.

The bearings were designed to provide a low-friction support at all speeds and a means of measuring this bearing resistance accurately. To accomplish this, two hydrostatic concentric journal-bearing pairs were used to mount each disk. The central journal was fixed relative to the frame. The bearing for the fixed journal, which also acted as the journal for the outer bearing was free to rotate, but in operation it only rotated a few degrees, sufficient to indicate the friction torque applied to it by the rotating disk and external bearing. Since the intermediate member consisting of the internal bearing and external journal did not rotate, no frictional resistance existed in the internal bearing.

The disks were pressed together in peripheral contact by suspending one disk assembly as a pendulum and counterbalancing to produce the desired load between the disks, Fig. 2. Lubricant was supplied at ambient temperature to the contact region by gravity from a reservoir located above the disks. The line OE is maintained vertical. Since each disk is subjected to a motor couple, a bearing-friction couple, a windage couple, and a resultant disk load and its reaction at the bearing, the resultant disk force of A on B is, Fig. 2

$$F = \frac{Tx - Wy}{OD} = \frac{Tx - Wy}{(J - R_b \tan \phi) \cos \phi} \quad \dots \dots [1]$$

However the angle ϕ is small (less than 3 deg) and therefore

$$F = \frac{Tx - Wy}{J} \text{ (with less than 2 per cent error)} \dots [2]$$

From Fig. 2 and Fig. 3, which is a free-body diagram of disk B the angle of inclination

$$\phi = \sin \phi = \frac{L_A + L_B}{R_A + R_B} \text{ (for small angles)} \dots \dots [3]$$

and the offset

$$h = \frac{R_B L_A - R_A L_B}{R_A + R_B} \text{ (except for a small difference)} \dots [4]$$
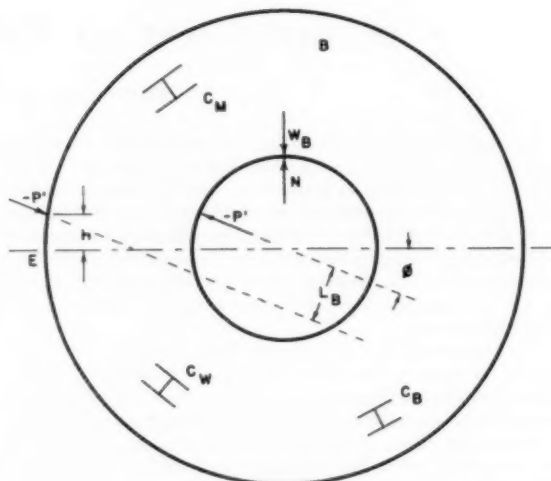
This paper includes a report of results obtained in a limited number of tests using the apparatus described. Each test was made using a copious amount of lubricant (at ambient temperature) supplied to the contact region, a constant counterbalance load, and operating one disk at a constant angular speed. A series of tests was made at constant load, constant angular velocity on one disk, and approximately constant viscosity lubrication (ambient temperature). The angular velocity of the other disk was varied from approximately 100 to 1200 rpm and the speed of the constant-speed disk was changed from test to test. Tests also were made using different constant loads and different viscosity lubricants.

RESULTS

Angularity Correlation. In one special case of curved surfaces in contact, viz., journal bearings, frictional resistance has been correlated with the dimensionless parameter ZN/P where Z is the viscosity of the lubricant, N is the shaft speed, and P is the applied load on the journal divided by the projected area of the bearing. It therefore seems probable that the frictional resistance, in the more general case reported herein, would be a function of this same or a similar parameter. It also seems reasonable that some other parameters are involved because of the differences in the more general situation and the special case of concentric cylindrical surfaces.

In attempting to correlate the data obtained, the ZN/P parameter was changed to fit the more general case. The N was replaced by the relative velocity between the theoretical points of contact and the P was changed to the load per unit length of line
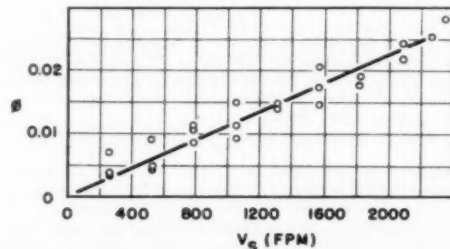
FIG. 4 ANGLE OF INCLINATION OF LINE OF ACTION OF RESULTANT
FORCE AS A FUNCTION OF SLIDING VELOCITY

contact giving ZV_s/P. The angle ϕ was first plotted against the sliding velocity V_s for constant disk load and constant-viscosity lubrication, Fig. 4. The data from this test and similar tests for different loads and viscosities were then plotted in Fig. 5 (ϕ versus ZV_s/P). The sliding velocity V_s still does not uniquely define the velocity conditions of contact because the same sliding velocity can be obtained in an infinite number of ways. In order, therefore, to describe a unique condition some other velocity parameter is needed. In developing such a relative-motion parameter it seemed reasonable and desirable to compare the relative motion with that existing when "pure rolling" occurs, since the theoretical condition of pure rolling is easily defined and readily recognized. Further it seems probable that the frictional resistance between the two contacting surfaces is in some way a function of the proximity of the contact conditions to those of pure rolling.

In order to develop a scale for comparison, let n be the ratio of the numerically larger to the smaller sweep velocities of the two bodies where the sweep velocity is defined as the velocity of the point of contact over the surface.[5] This limits the value of n to be $-1 > n > +1$. The criterion chosen is

$$S = [(1 - n)/n]^2 \quad\ldots\ldots\ldots\ldots\ldots\ldots [5]$$

Fig. 6 is a graph of this equation showing the range of values of S and n. For pure rolling $n = 1$ and $S = 0$. The quantity S is called the "rolling-contact proximity factor." Fig. 7 is another way of showing the rolling-contact proximity scale. Finally it was found that the data could be fairly well represented by the following empirical equation

$$\phi = 4.3 \left[\frac{ZV_s}{P}\right]^{2/3} S^{1/3} \quad\ldots\ldots\ldots\ldots\ldots [6]$$

Fig. 8 shows the agreement between the values of ϕ obtained experimentally and Equation [6].

Offset Correlation. The displacement of the line of action of the resultant force between the disks seemed to be independent of the sliding velocity. It again seems probable that the offset of the line of action of the resultant force between the contacting bodies is a function of a dimensionless parameter involving load, viscosity, and velocity. Since the offset does not seem to be a function of the difference of the sweep velocities (sliding velocity) and since it is desirable to define the contact situation by a unique velocity, the sum of the sweep velocities was considered. The pressures developed in the load supporting film will be a function of how rapidly the lubricant is drawn into the contact region and how rapidly the lubricant is squeezed out of this region. In addition, if hysteresis losses occur in the solid members, the line of action will move in the direction of the sweep velocity on each member. Hysteresis losses increase as the sweep velocities increase. The dimensionless parameter tried was ZV_Σ/P where Z and P are the viscosity and load respectively, and V_Σ is the sum of the sweep velocities.

It was found that the experimental data could be represented fairly well by the following empirical relationship

$$h_r/R = 0.24 \left[\frac{ZV_\Sigma}{P}\right]^{1/3} \quad\ldots\ldots\ldots\ldots\ldots [7]$$

where h_r is the offset of the line of action of the resultant force between the disks when the ratio n is unity (pure rolling), R is the relative radius of curvature of the disks, and V_Σ is the sum of the sweep velocities. Fig. 9 shows the agreement between the ex-

[5] The term "sweep velocity" was introduced by W. A. Tuplin in "Gear Tooth Lubrication," *Machine Design*, Penton Publishing Company, August, 1954, pp. 125–131.
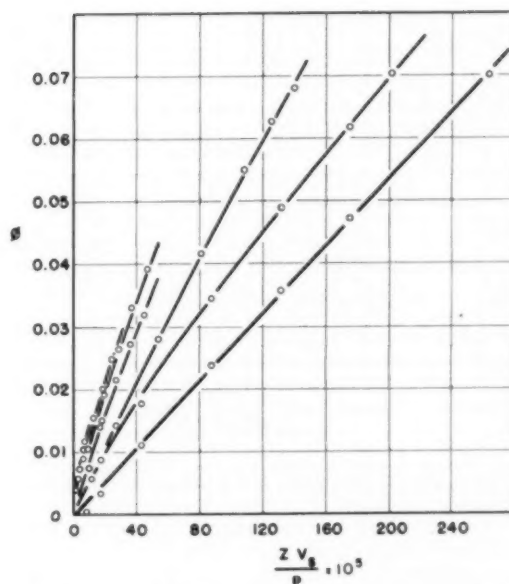
FIG. 5   ANGLE OF INCLINATION ϕ AS A FUNCTION OF ZV_s/P
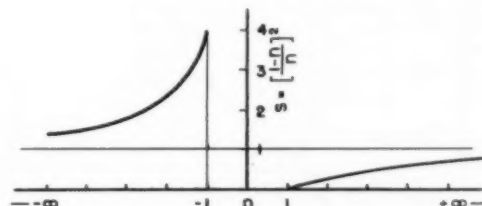
FIG. 6   ROLLING-CONTACT PROXIMITY FACTOR S

FIG. 7   ROLLING-CONTACT PROXIMITY SCALE
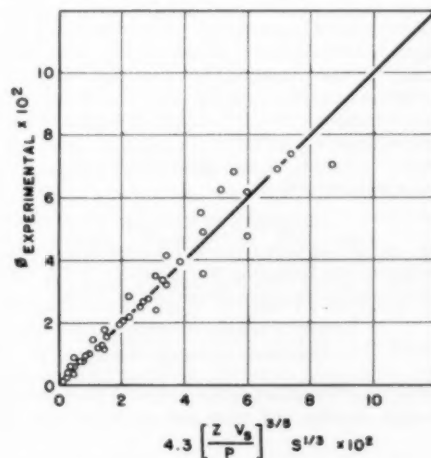
FIG. 8 $\phi_{experimental}$ VERSUS 4.3 $(ZV_s/P)^{2/3} S^{1/3}$
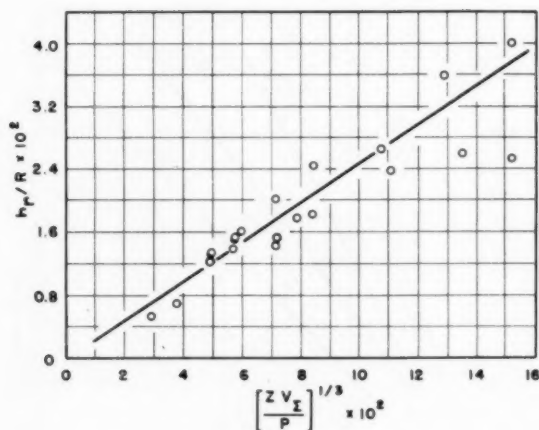
FIG. 9 OFFSET h_r FOR PURE ROLLING CONTACT DIVIDED BY RELATIVE RADIUS OF CURVATURE AS A FUNCTION OF $(ZV_\Sigma/P)^{1/3}$
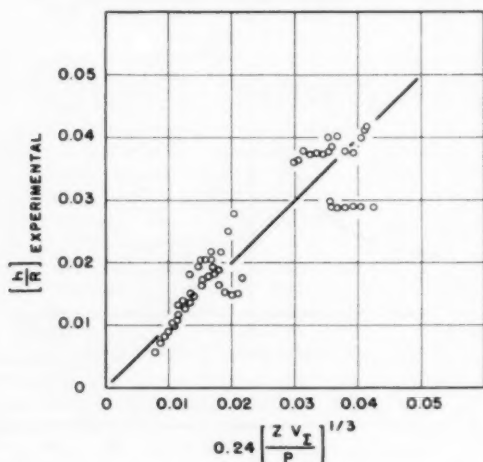
FIG. 10 $(h/R)_{experimental}$ VERSUS $0.24 (ZV_\Sigma/P)^{1/3}$

perimental results and Equation [7]. The offset h (for any value of n) seemed to be independent of the rolling-contact proximity factor S, and the same equation (Equation [7]) was found to fit the data fairly well as shown in Fig. 10. The correlation for h, Fig. 10, is not so good as that for h_r, Fig. 9, indicating that perhaps some other variable is important.

The results presented in this paper are based on a limited number of experimental data.

DISCUSSION OF RESULTS

Merritt (op. cit.) obtained results for a coefficient of friction which decreased with increasing sliding velocity. His coefficient was obtained from energy measurements and was defined as

$$\mu = \frac{\text{Torque} \times \text{angular speed of one disk}}{\text{Disk load} \times \text{sliding velocity}} \quad \ldots \ldots [8]$$

The loads used by Merritt are comparable with those encountered in worm-gear practice and were such as to produce contact
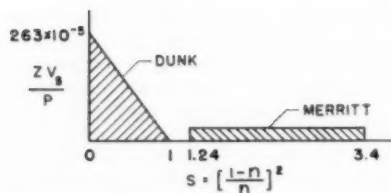
FIG. 11 REGION OF TEST DATA COMPARED WITH THAT OF MERRITT
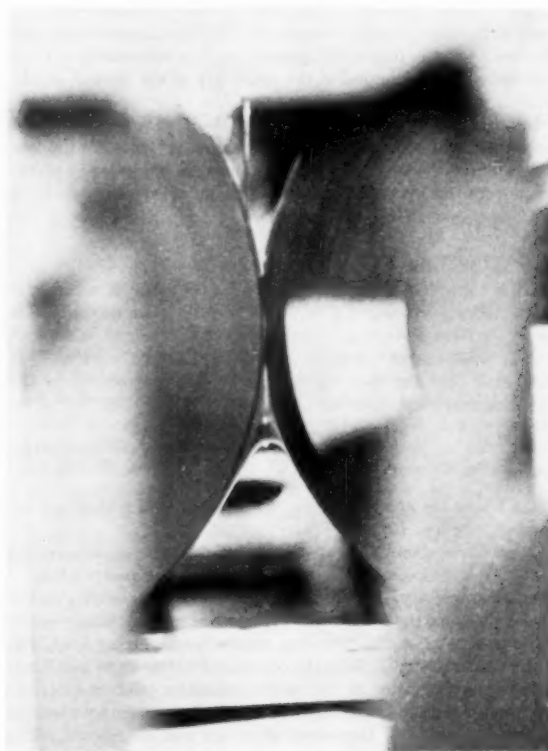
FIG. 12 OIL ARC ON RECESS SIDE OF CONTACT REGION

stresses of approximately 70,000 psi calculated by the use of Hertz' equations. This stress is considerably above the elastic limit of bronze (of which at least one disk was made) and so considerable inelastic deformation of the surfaces probably occurred, since metal-to-metal contact is probable at this load. Under these conditions, then, the frictional resistance would be a function of the materials, the oiliness of the lubricants, and the sliding velocity, among other factors. Merritt's results, then, seem to be in agreement with those obtained for journal bearings in the region of boundary lubrication, since both applications show a decreasing coefficient of friction for increasing sliding velocities.

Most of the tests reported on herein were made with contact stresses which would have been approximately 4200 psi if metal-to-metal contact occurred (static). The elastic limit for the weaker material is 12,000 psi. The region of the test data reported on herein is compared with that of Merritt in Fig. 11.

In the region investigated by the authors the frictional resistance increases with increasing sliding velocity, as it does in journal bearings for a constant load and viscosity and where thick-film lubrication is maintained. In the tests reported on herein thick-film lubrication was shown to exist by using an electrical continuity test between the disks. Fig. 12 is a photograph of the oil arc present between the disks while they are rotating in contact.

Conclusions

From the analysis of the contact conditions between curved surfaces and the results obtained in this investigation, it appears necessary and sufficient to describe the frictional resistance between two bodies resulting from their contact, by two quantities which locate the position of the line of action of the resultant force between the bodies. The angle of inclination ϕ and the offset h of the line of action of the resultant force between the contacting surfaces is a suitable pair of quantities.

Of the various velocities involved, it is the sliding velocity that is of primary importance in determining the inclination of the force line.

It is the sum of the sweep velocities that is of primary importance in determining the offset of the force line.

The results were not conclusive with respect to the dependence of the angle of inclination or the offset on the rolling-contact proximity factor.

Discussion

J. H. Billings.[6] The authors have investigated in a field of great scientific and practical interest. As long as gears and cams transmit and wheels roll, there will be need for exact knowledge on the mechanics of sliding and rolling.

For discussion of the points raised, the writer finds it necessary to consider the total frictional resistance resulting from sliding and rolling, for which the authors have no symbol. Call it f. This friction force imposed on disk B, Fig. 2, will act normal to the line of centers and through the theoretical contact point. Then the coefficient of friction equals

$$\frac{\text{Friction force}}{\text{Normal force}} = \frac{f}{W} = \tan \phi = \phi \text{ (nearly)}$$

The resultant load P, between the disks, will always be at the angle ϕ with the line of centers since the resultant of a coplanar force and a couple is parallel to the force. This means that the authors' ϕ is, for small angles, a close measure of the coefficient of friction.

With this definition of ϕ, Fig. 5 of the paper is revealing. Ordinate $\phi = f/W$ and W is closely proportional to P in the abscissa. Experimental data giving a straight line pointing through the origin tells us that friction force f is directly proportional to shearing velocity when viscosity Z is constant. This is an accepted law of fluid mechanics. The conclusion from the authors' data then is that the load between the disks is being completely carried hydrodynamically by lubricant film. This is substantiated by the authors' statement that electrical transmission between the disks was found to be absent.

The authors' statement that they were unable to establish a useful relationship between performance and proximity to the condition of pure rolling is interesting and I think this result was to be expected. Pure rolling represents the most ideal condition for large hydrodynamic load-carrying capacity between surfaces mutually convex. The two surfaces join forces in squeezing fluid into the pressure region.

[6] Head of Mechanical Engineering Department, Drexel Institute of Technology, Philadelphia, Pa. Fellow ASME.

Indeed, proximity to the condition most remote from pure rolling would be a more promising criterion. Suppose the two surfaces move past the theoretical contact point at the same speed in opposite sense. Now, one surface tends to force fluid into the pressure region at the same rate that the other surface carries it out and the net pumping effect is zero—likewise the hydrodynamic load-carrying capacity.

It is worth noting that, in applying this principle to the general case, surface velocities must be measured relative to the contact point since, as in cams and gearing, the contact point itself may be moving.

I suggest to the authors that they run some tests on this equipment with conditions remote from pure rolling. I would expect such data plotted in Fig. 5 of the paper to give curves with decided hooks. I think we are not yet ready to discard the coefficient of friction as the most useful measure of performance where transmission is by direct sliding and rolling contact.

The authors are to be especially congratulated on the ingenuity displayed in the design of this equipment.

H. E. Merritt.[7] The writer wishes to thank the authors for their references to his 1935 publication, and compliment them on what he believes to be the first piece of experimental research on disk friction since his purely pragmatic approach so long ago.

From the point of view of pure physics, the authors may be correct in expressing resistance to motion in terms of the inclination and offset of the resultant line of force. Nevertheless, in the expectation that data obtained from researches of this kind may in due course find application to engineering practice, would it not be possible to express the effects of inclination and offset individually in terms of the ratio of tangential resistance to normally applied load, i.e., the usual concept of a coefficient of friction?

It would have been helpful if the authors had presented more details of the experimental facts and observations proper, viz., disk dimensions and materials, nature of the lubricant, and the speeds, loads, and torques. This would have assisted others interested in the subject to undertake their own analysis of, or speculations about, the results obtained.

A great deal more work of this kind is needed. There is a serious dearth of information required for the practical problems of estimating gear-tooth efficiency; and more fundamentally, it is probable that such data would throw more light on the behavior of gear materials, with particular reference to the phenomena of scuffing and pitting, both of which must be influenced by the tangential stresses and the temperatures in the surface layers.

It is therefore to be hoped that more work will be done, and at much higher pressures. The authors quote their nominal pressure as 4200 psi, which is far below operating gear-tooth pressures. These pressures range from 50,000 to 300,000 psi; but the writer prefers the criterion S_c = load per inch of line contact per inch of relative radius of curvature. In this pressure range, it seems probable that the viscosity effects investigated by the authors become completely submerged by frictional phenomena of an entirely different kind. The nature of these phenomena remains unexplored but the profound effect of the materials and the differences in frictional values obtained with lubricants of apparently similar viscosity present a challenge which it is hoped will be accepted by research workers.

On a point of detail, the authors have apparently misinterpreted the expression for coefficient of friction given in my cited paper; and this does not lead to an infinite value at zero sliding velocity. I have given the value of 0.15 determined experimentally.

[7] Lion Hill, Claverdon, Warwick, England.

Authors' Closure

The authors wish to thank Professor Billings and Dr. Merritt for their discussions which are valuable contributions to the paper.

The actual contact region of two curved surfaces under load in the presence or absence of a lubricant is not a point and hence the resultant force cannot correctly be resolved into two perpendicular components at the theoretical point of contact as Professor Billings suggests. The resultant force does not pass through this point.

Dr. Merritt is quite correct in pointing out that his expression for coefficient of friction does not necessarily lead to an infinite value at zero sliding velocity. It does, however, lead to an infinite value under rolling conditions, where the energy loss is not zero but the sliding velocity is zero.

Comparison of Semi-Empirical Solutions for Crack Propagation With Experiments

By J. FRISCH,[1] BERKELEY, CALIF.

The crack-propagation characteristics of two aluminum alloys and several steels have been investigated in sheet materials under uniaxial tension. Empirical formulas based on ultimate strength and initial crack length were found to be a sufficiently effective method for establishing the relative merit of a material in its resistance to crack propagation. Comparison of experimental data with criteria obtained from the energy-release rate or stress-concentration theory shows that neither method yields reliable material constants.

NOMENCLATURE

The following nomenclature is used in the paper:

B = specimen width, in.
E = modulus of elasticity, ksi
E_u = secant modulus at ultimate tensile strength, ksi
K_t = theoretical stress-concentration factor
K_n = Neuber "engineering" stress-concentration factor
K_u = stress-concentration factor for ultimate tensile strength
L = specimen length, in.
P = maximum applied load, lb
r = radius of curvature at crack tip, in.
r' = effective radius of curvature, in.
t = specimen thickness, in.
U = strain energy, in-lb
W = work for crack propagation, in-lb
dW/dA = fracture work rate for crack propagation, in-lb/sq in.
x_0 = initial crack length, in.
x = crack length at maximum load, in.
σ = maximum gross-area stress, psi
σ_u = ultimate tensile strength, psi
η = plasticity modulus E_u/E
ksi = 1000 psi

INTRODUCTION

The sudden fracture and resulting catastrophic failures of various engineering structures such as ships, aircraft, tanks, pipelines, and the like, have over many years been the subject of an exceedingly large number of investigations to study their causes and prevent their recurrence. A survey of the innumerable publications in this field shows the need for further studies of crack propagation in materials alone. A recent evaluation of the knowledge on the subject by Drucker, et al. (1)[2] indicates the divergence of opinions as well as the difficulties in finding practical solutions. In the realization that a panacea for sudden fracture

does not exist, many aspects of the design of structures capable of withstanding catastrophic failure even though considerable damage has occurred, i.e., concepts of fail-safe design, have been investigated by Conway (2), Sorensen (3), Wells (4), and many others.

The determination of the load-carrying capacity of designs of complex configurations assumes a knowledge of the ultimate strength and other mechanical properties of the materials used in the manufacture of such structures. It is, therefore, reasonable to assume that the relative tear resistance of various materials must be known before the strength of damaged structures, made from such materials, can be ascertained. Two methods of finding such a criterion are (a) the energy balance between strain energy and work necessary to cause fracture as postulated by Griffith (5) and used in modified form by Irwin and Kies (6, 7) and (b) the use of a stress-concentration factor described by McEvily, Illg, and Hardrath (9). The attempt has been made in both methods to find a material constant. In the case of the energy solution, the fracture work rate dW/dA necessary to extend a crack is thought to be fairly constant for a material and, for the stress-concentration solution, an effective crack-tip radius r' in place of the geometric radius is considered as a similarly constant value. However, an evaluation of test data shows that values of dW/dA and r' for the materials under consideration varied by as much as 3.35 to 1 and 5.50 to 1, respectively. With such large variations it was considered desirable to write empirical equations which can be used for more reliable estimates of tear resistance. The experimental results of this investigation also show that the size of the test specimens greatly influences the values of both the fracture work rate and the effective radius.

The materials used were aluminum alloys 2024-T3 Alclad and 7075-T6 Alclad, as well as stainless steels 17-7PH, 301 fully hard, 302 annealed, and AM350. Although considerable data are available on various steels, the materials used here were selected for their current use as well as future application in high-temperature designs. The assumptions made in both theories are such that the critical stress becomes inversely proportional to the square root of the crack length. This was found to be the case for ship steel as reported by Felbeck and Orowan (11) as well as in some of the materials reported here. However, the empirical relationships for the materials shown in Table 1 indicate that this is not necessarily the case for either all materials or, for that matter, for different size panels of the same material.

THEORY

The energy principle for self-propagation of a crack as stated by Griffith (5) represents one of the earliest efforts toward a solution for fracture problems. It states that if dW represents the free energy necessary for increasing the length x_0 of a crack such as the one shown in Fig. 1, and $-dU$ the elastic energy released simultaneously in the specimen, the critical crack length above which spontaneous propagation takes place can be determined by the equation

$$dW = -dU \quad\quad\quad\quad\quad [1]$$

To apply the principle to ductile materials, Irwin and Kies (6, 7) took into account the rate of work necessary to extend the frac-

TABLE 1 SUMMARY OF EXPERIMENTAL DATA AND MATERIAL CONSTANTS

1	2	3	4	5	6	7	8	9	10
Material	Spec. No.	Width B, in.	Thickness t, in.	Initial crack length x_0, in.	Crack length at max load x, in.	Max gross area stress σ, ksi	$\frac{dW}{dA}$ in-lb sq in.	r', in.	Empirical equation $\sigma = K\sigma_u/x_0{}^n$
2024-T3 Alclad aluminum; $\sigma_u = 62$ ksi $\eta = 0.034$	1	9	0.091	0.75	1.01	46.4	337	0.0153	$\sigma = 0.88\,\sigma_u/x_0{}^{0.576}$
	2	9	0.091	1.84	2.36	38.4	592	0.0113	
	3	9	0.091	3.57	4.36	26.3	658	0.0045	
	4	20	0.091	1.79	3.12	44.0	960	0.0247	
	5	20	0.091	3.58	4.82	35.8	1040	0.0155	$\sigma = \sigma_u/x_0{}^{0.560}$
	6	20	0.091	5.40	6.80	26.5	865	0.0069	
7075-T6 Alclad aluminum; $\sigma_u = 72$ ksi $\eta = 0.071$	7	9	0.091	0.76	1.17	44.6	373	0.0203	$\sigma = 0.63\,\sigma_u/x_0{}^{0.448}$
	8	9	0.091	1.92	2.27	36.9	533	0.0214	
	9	20	0.091	1.81	2.40	32.7	317	0.0126	
	10	20	0.091	3.55	4.15	24.2	397	0.0092	$\sigma = 0.63\,\sigma_u/x_0{}^{0.431}$
	11	20	0.091	5.34	6.07	19.0	384	0.0069	
17-7PH stainless steel; $\sigma_u = 209$ ksi $\eta = 0.0727$	12	9	0.085	1.27	1.47	65.6	352	0.0007	$\sigma = 0.36\,\sigma_u/x_0{}^{0.330}$
	13	9	0.085	2.94	3.10	51.5	514	0.0008	
	14	20	0.085	2.93	2.93	61.1	600	0.0013	
	15	20	0.085	5.74	7.00	52.0	1180	0.0017	$\sigma = 0.36\,\sigma_u/x_0{}^{0.138}$
302 annealed stainless steel; $\sigma_u = 91$ ksi $\eta = 0.0552$	16	9	0.081	1.51	1.82	46.2	210	0.0024	$\sigma = 0.665\,\sigma_u/x_0{}^{0.437}$
	17	9	0.081	3.16	3.25	36.3	274	0.0021	
	18	20	0.081	3.02	3.85	41.0	366	0.0031	
	19	20	0.081	4.95	6.05	35.8	464	0.0031	$\sigma = 0.665\,\sigma_u/x_0{}^{0.290}$
301 fully hard stainless steel; $\sigma_u = 180$ ksi $\eta = 0.0776$	20	9	0.081	1.38	1.82	133.5	1745	0.0312	$\sigma = \sigma_u/x_0{}^{0.500}$
	21	9	0.081	2.88	3.25	97.3	1948	0.0114	
	22	20	0.081	3.38	4.35	115.0	3270	0.0341	
	23	20	0.081	5.79	7.35	86.2	3510	0.0152	$\sigma = \sigma_u/x_0{}^{0.328}$
AM350 stainless steel; $\sigma_u = 200$ ksi $\eta = 0.0835$	24	9	0.066	1.40	1.70	150.2	2170	0.0445	$\sigma = \sigma_u/x_0{}^{0.540}$
	25	9	0.065	2.85	3.15	108.6	2310	0.0140	
	26	20	0.064	2.97	3.50	136.2	3620	0.0473	
	27	20	0.066	4.72	7.40	111.0	4040	0.0256	$\sigma = \sigma_u/x_0{}^{0.296}$

ture area. Considering the simultaneous energy-release rate, Equation [1] at critical crack length becomes

$$\frac{dW}{dA} = -\frac{dU}{dA} \qquad\qquad [2]$$

By assuming the released energy to have been contained in an ellipse having crack length x as its minor axis and $2x$ as its major axis the total energy is

$$U = \frac{\pi x^2 t}{2}\frac{\sigma^2}{2E} \qquad\qquad [3]$$

The fracture work rate then is

$$\frac{dW}{dA} = \sigma^2\frac{\pi x}{2E} \qquad\qquad [4]$$

or the critical stress

$$\sigma = \left(\frac{2E\,dW/dA}{\pi x}\right)^{1/2} \qquad\qquad [5]$$

Taking into account the boundary effect due to finite width B of a test specimen, corrections for x/B ratios have been computed (6, 7).

Equation [4] indicates that a relatively constant fracture work rate should be obtained for any given material. However, as has been reported (7) and as shown in Table 1, variations of dW/dA of as much as 3.35 to 1 in relatively small and large specimens of a material have been encountered.

Consideration of a central crack in a flat sheet, such as the one shown in Fig. 1, as an elongated ellipse has led to stress-concentration analyses (9) of the problem. The theoretical stress-concentration factor K, for such an ellipse is given as
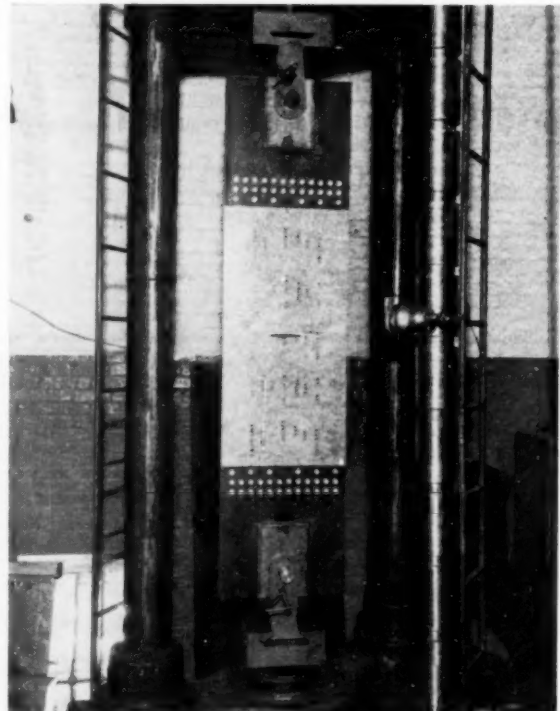

FIG. 1 TYPICAL CRACK-PROPAGATION SPECIMEN IN UNIAXIAL TENSION MOUNTED IN 400,000-LB TENSILE TESTING MACHINE (L/B RATIO 2)

$$K_t = 1 + 2 \left(\frac{x_0}{2r}\right)^{1/2} \dots [6]$$

Neuber's (10) modified engineering stress-concentration factor was shown to be (9)

$$K_n = 1 + \frac{1}{2}(K_t - 1) \dots [7]$$

Using the plasticity modulus η the ultimate strength stress-concentration factor K_u becomes

$$K_u = 1 + \eta(K_n - 1) \dots [8]$$

From substitutions of Equations [6] and [7] as well as the use of an effective radius r' rather than the actual crack-tip radius r

$$K_u = 1 + \eta \left(\frac{x_0}{2r'}\right)^{1/2} \dots [9]$$

and the critical stress in terms of ultimate strength is therefore

$$\sigma = \frac{\sigma_u}{1 + \eta \left(\frac{x_0}{2r'}\right)^{1/2}} \dots [10]$$

Equation [10] indicates that a relatively constant value of r' should exist for a given material. However, for the materials tested, values for the effective radius varied by as much as 5.50 to 1.

Empirical Equation. A comparison of the critical gross-area stress with the ultimate strength for various crack lengths shows that the following empirical relationship exists for the materials presented here

$$\sigma = \frac{K\sigma_u}{x_0{}^n} \dots [11]$$

where K is a constant and $n < 1$.

Test results show that the value of $n = 0.5$ based on Equations [5] and [10] does not remain constant but varies with materials and specimen sizes.

Experimental Technique

The specimens used in the investigation were 9×20 in. and 20×40 in. in size with thicknesses as shown in Table 1. The two aluminum alloys under consideration were 2024-T3 Alclad and 7075-T6 Alclad sheet materials. The 301 fully hard and 302 annealed stainless-steel specimens were tested with the heat-treatment as indicated while the 17-7PH[3] and AM350[4] stainless-steel specimens were subjected to heat-treatments before initiation of the artificial crack and subsequent propagation tests.

The initial crack for the aluminum specimens was produced with a 0.008-in. jeweler's saw. It was found that the crack-propagation data obtained from such saw cuts were equivalent to those determined from fatigue cracks. However, in the case of the stainless steels, the inefficiency of the jeweler's saw necessitated the forming of the initial crack by cycling the specimens in a tension-tension fatigue machine until the desired crack length had been attained. The crack-propagation tests were conducted in a 400,000-lb tensile testing machine with the load applied through friction grips for the 9-in. specimens and jig plates for the 20-in. specimens. A typical specimen and test setup is

[3] From annealed condition, heated to 1400 F for $1^{1}/_{2}$ hr and air-quenched. Within 1 hr after quench, cooled below 60 F for 30 min. Tempered at 1050 F for $1^{1}/_{2}$ hr.

[4] From annealed condition, subzero cooled to -100 F for 2 hr and tempered at 750 F for $2^{1}/_{2}$ hr.

shown in Fig. 1. The crack extensions were measured with a scale of 0.01 in. least count at proper intervals as the load was increased to maximum, corresponding to critical crack length, after which the crack becomes self-propagating.

However, because of the design of tensile testing machines the crack growth after the critical length had been reached is associated with load drops in the machine. The gross-area stress based on the maximum attained load and the corresponding crack length x as well as the initial crack length x_0 were therefore considered as the critical conditions for subsequent evaluations.

Results and Discussion

The influence of the boundary effect, due to finite width B of the specimen, on the critical failure stress σ is demonstrated in Figs. 2 and 3(a). It was found that the 20-in-wide specimens will fail at a consistently higher stress level than the 9-in. ones. As shown in Fig. 6, the 35-in-wide specimens of 2024-T3 aluminum failed at about the same stress level as the 20-in. specimens. It appears therefore that the latter width is sufficient for obtaining data unaffected by specimen width. Consequently, the subsequent tests were limited to a maximum width of 20 in. The erroneous conclusions which could be derived if only small specimens, 9 in. wide or less, were used as an indication of relative crack propagation strength, is best illustrated by a comparison of the two aluminum alloys shown in Figs. 3(a) and (b). The better crack-propagation characteristics as well as the considerably higher values of dW/dA did not become apparent until wider specimens, 20 in. or more, were used.

Values of the fracture work rate dW/dA have been calculated from the test results and are plotted in Figs. 3(b) and 4 as functions of specimen width B. The ranges of dW/dA for each
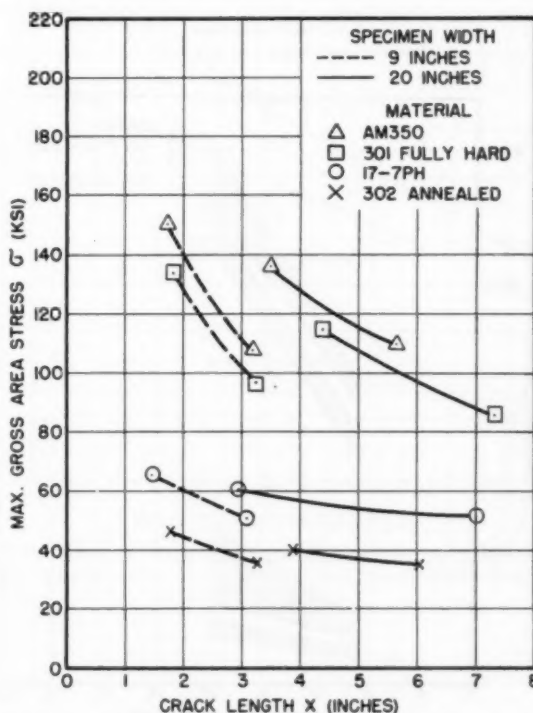

Fig. 2 Maximum Gross-Area Stress Before Self-Propagation of Crack as a Function of Crack Length at That Stress Level (9 and 20-in. stainless-steel sheet s.)
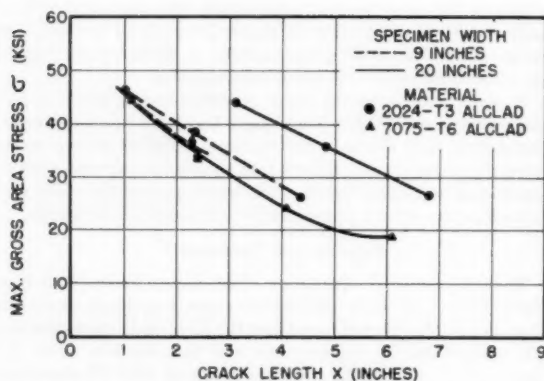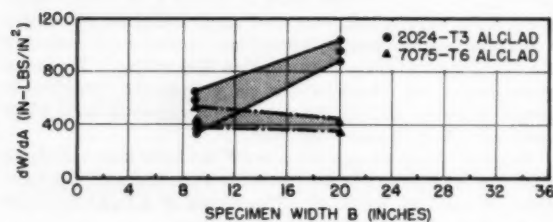
(b) Fracture Work Rate as a Function of Specimen Width B for Aluminum Sheets
(Based on crack length x at maximum load.)
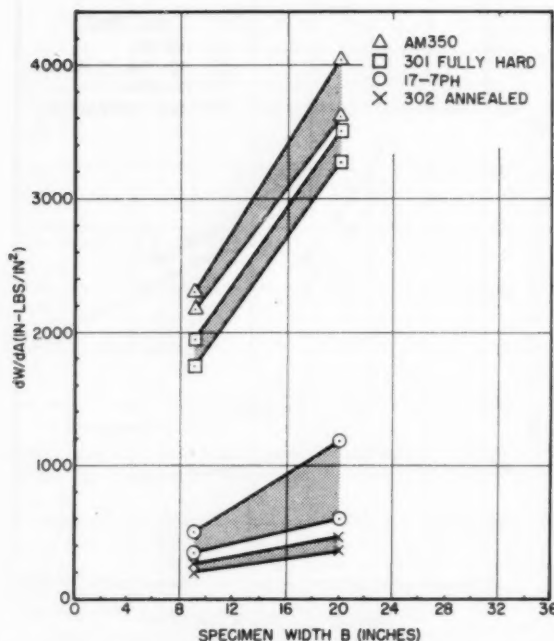
Fig. 4 Fracture Work Rate as a Function of Specimen Width B for Stainless-Steel Sheets
(Based on crack length x at maximum load.)

material are shown in Table 1 and the ratios of maximum to minimum values in Table 2. While the materials can be arranged in a qualitative manner with respect to their resistance to crack propagation, the variations of fracture work rates for any one material by as much as 3.35 precludes accurate estimates of the critical-failure stresses for any particular crack length.

Table 2 Ratios of Maximum to Minimum Values of Fracture Work Rates and Effective Stress-Concentration Radii

Material	$\dfrac{(dW/dA)\ \text{max}}{(dW/dA)\ \text{min}}$	$\dfrac{(r')\ \text{max}}{(r')\ \text{min}}$
2024-T3 Alc. aluminum	3.08	5.50
7075-T6 Alc. aluminum	1.62	3.10
17-7PH stainless steel	3.35	2.37
302 ann. stainless steel	2.21	1.50
301 FH stainless steel	2.01	2.99
AM350 stainless steel	1.86	3.38

The effective radii r' for calculating the ultimate strength-stress concentration factors were determined from the test results and are plotted as functions of panel width B in Fig. 5. Although the notch sensitivity of the material has been taken into account through use of the plasticity modulus $\eta = E_s/E$, the range of these effective radii is considerable, as evidenced by the ratios of up to 5.50 shown in Table 2, for maximum to minimum values of r'. The value of such an effective radius hinges on the accurate determination of the secant modulus at ultimate tensile strength. Since this modulus for materials with high percentage elongation near the ultimate tensile strength is difficult to determine, it may very well become an additional source of error in the calculation of effective crack-tip radii.

In the absence of either a reasonable constant fracture work rate or effective radius for the materials tested, a comparison of the critical crack-propagation stress σ with the ultimate tensile strength σ_u has been made for various initial crack lengths x_0 larger than 1 in. The ratios σ/σ_u versus x_0 for the test results of 9 and 20-in. panels are plotted logarithmically in Figs. 6 and 7. From Equation [11] the relationship between the critical stress and the initial crack length was obtained and is listed in Table 1. Taking experimental errors into account, it can be seen that σ is approximately inverse proportional to $\sqrt{x_0}$ as predicted in Equations [5] and [10] for the two aluminum alloys as well as the smaller specimens of AM 350 and 301 FH stainless steels. However, for the remaining materials reported here as well as different panel widths the exponents of x_0 were found to be less than 0.5.

A varying strain-energy distribution over the surface of specimens identical to those used in this investigation has been reported by Liu and Carpenter (8). Considering in addition the plastically deformed area surrounding the crack tip before propagation, it can be seen that these factors must be included in a reliable criterion for measuring the crack-propagation resistance of materials. While the foregoing considerations remain subjects for further research, the empirically determined equations presented here may be considered a sufficient method of evaluating materials.

Conclusions

1 The experimentally determined fracture work rates and effective stress-concentration radii, which according to theory should remain relatively constant for each material, were found to vary by as much as 3.35 and 5.50, respectively, for the aluminum alloys and stainless steels investigated.

2 Without proper corrections for size effect and similar variables, experimentally determined values of dW/dA and r' cannot be used for estimating critical crack-propagation criteria.

Fig. 5(a)

Fig. 5(b)

FIG. 5 (a–f) EFFECTIVE STRESS-CONCENTRATION RADIUS r' AS A FUNCTION OF SPECIMEN WIDTH B FOR ALUMINUM AND STAINLESS STEEL SHEETS
(Based on initial crack length x_0.)
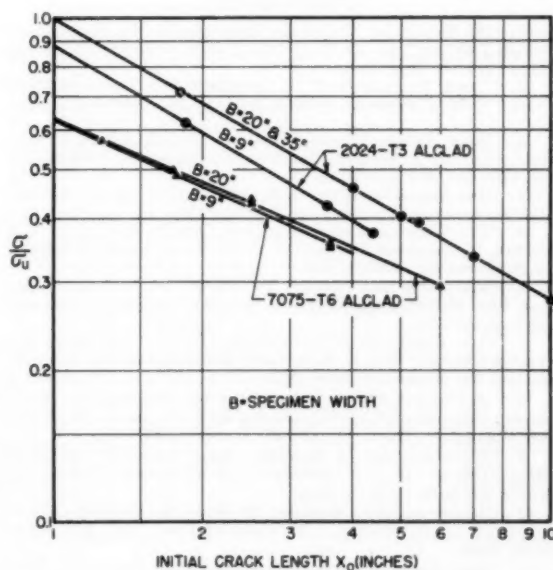
FIG. 6 RATIO OF MAXIMUM GROSS-AREA STRESS TO ULTIMATE TENSILE STRENGTH AS A FUNCTION OF INITIAL CRACK LENGTH x_0 FOR ALUMINUM SHEETS
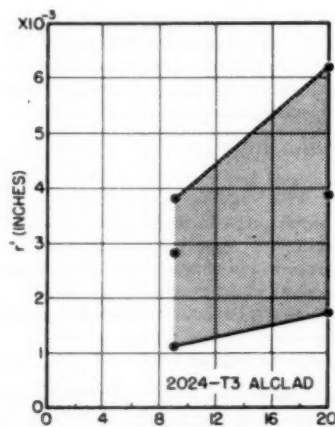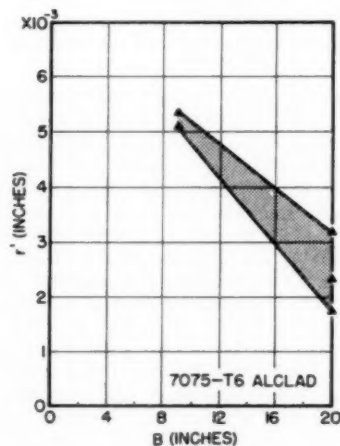
3 The investigation showed that the critical-failure stress is not necessarily inverse proportional to $(x_0)^{1/3}$ as assumed from theory.

4 Empirical equations for the maximum gross-area stress of centrally cracked flat sheets in uniaxial tension were found to be a sufficient expression for estimating the tear resistance of materials.

FIG. 7 RATIO OF MAXIMUM GROSS-AREA STRESS TO ULTIMATE TENSILE STRENGTH AS A FUNCTION OF INITIAL CRACK LENGTH x_0 FOR STAINLESS-STEEL SHEETS

BIBLIOGRAPHY

1 "An Evaluation of Current Knowledge of the Mechanics of Brittle Fracture," by D. C. Drucker, Ship Structure Committee Report No. 69, May, 1954.

2 "Crack Propagation Tests Safeguard New Lockheed Structures," by W. J. Conway, *Aviation Age*, vol. 26, December, 1956, p. 32.

3 "Some Design Considerations for Tear-Resistant Airplane Structures," by A. Sorensen, Institute of Aeronautical Sciences, Preprint No. 618, January 23, 1956.

4 "The Mechanics of Notch Brittle Fracture," by A. A. Wells, *Welding Research* (London), vol. 7, April, 1953, pp. 34r–56r.

5 "The Phenomena of Rupture and Flow in Solids," by A. A. Griffith, Philosophical Transactions of the Royal Society of London, England, vol. 221, 1920, p. 163.

6 "Fracturing and Fracture Dynamics," by G. R. Irwin and J. A. Kies, *Welding Journal*, vol. 31, Research Supplement, February, 1952, pp. 95s–100s.

7 "The Resistance of Materials to Fracture Propagation and Gunfire Damage," by J. A. Kies, Naval Research Laboratory Memorandum No. Report 594, May, 1956.

8 "A Study of Plastic Deformation and Fracturing by Strain Energy Distribution," by S. I. Liu and S. T. Carpenter, Ship Structure Committee Report No. SSC-38, December 20, 1950.

9 "Static Strength of Aluminum Alloy Specimens Containing Fatigue Cracks," by A. J. McEvily, Jr., W. Illg, and H. F. Hardrath, NACA TN 3816, October, 1956.

10 "Theory of Notch Stresses: Principles for Exact Stress Calculations," by H. Neuber, J. W. Edwards Bros., Inc., Ann Arbor, Mich., 1946.

11 "Experiments on Brittle Fracture of Steel Plates," by D. K. Felbeck and E. Orowan, *Welding Journal*, vol. 34, Research Supplement, 1955, pp. 570s–575s.

Discussion

F. A. McCLINTOCK.[5] In reference (3) of the paper, Sorenson points out that the tear resistance, or work per unit of crack area, not only varies with thickness, but varies in different ways for different materials. Does the author have any further data on the effect of thickness?

---

[5] Associate Professor of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Mass. Mem. ASME.

E. OROWAN.[6] The lack of agreement between the author's measurements and the equations assumed to represent the theoretical background is not surprising; in the case of ductile crack-propagation processes, Equations [1] and [2] are manifestly inapplicable. The corresponding correct equation is

$$dL = dW + dP + dU \qquad \qquad [12]$$

where

dL = work of external forces during a differential increase of crack length

dW = increase of free energy of crack walls (i.e., of their surface energy and of the plastic surface work[7])

dP = plastic work absorbed in bulk deformation around crack (excluding plastic surface work)

dU = increase of elastic energy of body

If the crack-propagation process is essentially brittle, the bulk plastic work dP vanishes; it is easy to prove then[8] that, at constant load

$$dL = 2\, dU \qquad \qquad [13]$$

and so Equation [12] is reduced to

$$dW = dU$$

which is the author's Equation [1], apart from the sign chosen for U. This is the case of the Griffith theory of purely brittle fracture, and of the writer's treatment of essentially brittle fracture in ductile steels.[7] However, if the fracture is essentially ductile, dP is not only not zero but usually becomes the dominating term on the right-hand side, and dW and dU are mostly negligible by comparison; naturally, Equation [13] and the consequent Equations [1] and [2] in the paper are then invalid. This is the case when the crack opens up by plastic deformation in its surroundings, not by the release of elastic strains. Thus in the common cup-and-cone fracture, frequently the crack opens up by an amount of the order of $1/_{16}$ in. in a round tensile specimen of some $1/_2$-in. diam. Since the elastic strains are of the order 10^{-3}, complete elastic-stress release above and be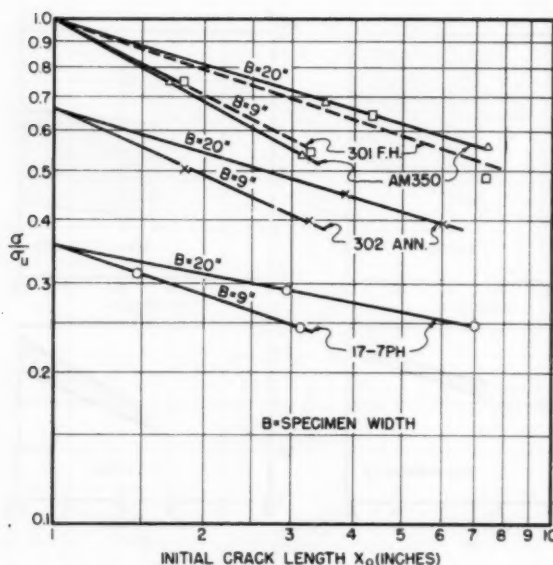low the crack would only open it up by about $1/_{1000}$ in.; the effect of elastic-strain release, even if it did occur on any substantial scale, would be negligible in comparison with the plastic opening up of the crack.

That Equations [1] to [5] in the paper are inapplicable to typically ductile crack propagation can be recognized directly. To take an extreme case, a cup-and-cone fracture in a material of given yield stress would take place in the same way and practically at the same load if the modulus of elasticity was infinitely high, since the crack-propagation process is one of plastic deformation in which elastic-strain release plays no significant part. On the other hand, Equation [5] which follows from Equations [1] and [2] would give an infinitely high strength (fracture stress) in this case. The use of the Griffith brittle-crack-propagation condition, Equation [1], for ductile crack propagation, therefore, is based on a misunderstanding; its cause was no doubt lack of accuracy in defining the meaning of the quantity W.

In any general energy equation, the work of the applied forces must be present as well as the energy items into which the work

---

[6] Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Mass.

[7] "Fundamentals of Brittle Behavior in Metals," in "Fatigue and Fracture of Metals," by E Orowan, Massachusetts Institute of Technology Symposium, 1950, Technology Press and John Wiley & Sons, New York, N. Y., 1952

[8] "Energy Criteria of Fracture," by E. Orowan, *Welding Journal, Research* (Welding Supplement), March, 1955.

is converted. This condition is not satisfied in Equation [1]. Even if W is assumed to include the bulk plastic work P, the work of the applied forces is missing from the equation. In the brittle case, crack propagation may occur at rigidly fixed grips, without any work done by the external forces*; in essentially ductile types of fracture, however, this is not so. Sometimes, if the specimen is extremely long compared with one of its transverse dimensions (e.g., if it is a long thin wire), elastic-strain release can lead to crack propagation. This, however, is the trivial case of a part of the specimen acting as a soft spring and effectively keeping the load constant during crack propagation in those parts which participate actively in the propagation process. It can be shown[9] that such processes are governed by a relationship between the *second* derivatives of W and U (not by Equation [1]), and the fracture stress cannot be obtained from the equation expressing the condition of rapid crack propagation. This, of course, is obvious; the strength of a specimen cannot be deduced from an equation containing the characteristics of an arbitrary elastic spring used in series with the specimen.

It is equally easy to show that any adaptation of the elastic stress-concentration factor given in Equation [6] to the ductile case by means of a fictitious "plasticity modulus" in the way given in Equations [8] to [10] is impossible. The elastic stress-concentration factor goes toward infinity with the square root of x_0/r; in the (two-dimensional) plastic case, however, the maximum possible local tensile stress (which occurs in the case of plane strain) converges toward $(1 + \pi/2)$ times the tensile yield stress, no matter how high x_0/r. The introduction of the factor η and the replacement of r by the quantity r' lose their physical meaning when plastic deformations spread beyond a small region around the crack tip, as must be the case in any genuine ductile crack-propagation process. In particular, there is no foundation for the assumption that "a relatively constant value of r' should exist for a given material." If two geometrically similar specimens, containing geometrically similar cracks are considered, and if one is 10 times larger linearly, not only the crack radius r but also the effective radius r' must be 10 times larger in it.

In spite of the lack of substantiality of its theoretical starting points, the author's investigation is valuable because it puts the spotlight upon a set of deeply ingrained misunderstandings about the theory of crack propagation which have been responsible for a considerable waste of time, money, and experimental ingenuity in recent years.

A. TILLES.[10] The author has presented an interesting experimental work, and brought into focus its relation to current theory about this complex subject.

His well-behaved experimental points show that, within the limited range of conditions of his tests, the crack-propagation characteristics of each of the several materials are predictable and dependent only on the length of the crack present and on properties of the material.

Which property of the material is the pertinent property is not ascertained, but is given a *de-facto* recognition in the numerical values ascertained for K and n.

The author points out that one of the logically appealing and widely studied material properties, the fracture work rate, dW/dA, does not fit the data as a constant, and is not the pertinent material property descriptive of its behavior in crack propagation.

Also, use of the effective crack radius r' is not applicable; that is, it is not applicable on the author's apparent assumption

[9] "Conditions of High-velocity Ductile Fracture," by E. Orowan, *Journal of Applied Physics*, vol. 26, 1955, p. 900.

[10] Robert E. Thomas & Associates, Engineers, 582 Market Street, San Francisco, Calif.

that the physical crack radius r present and the effective radius r' were really the same in all his samples.

Possibly this point deserves a little more attention than it was given in the necessarily limited presentation of the paper. What was the physical radius and what the effective radius in each case, at the beginning of the pull, and also at the moment of maximum stress? The author's estimate on these points would be of interest.

More detail than given on the comparative behavior of saw cuts and of fatigue cracks, with careful attention to the radii present, might be illuminating.

Does it not seem that a simple empirical relationship cognizant of material characteristics and crack lengths *only* must necessarily be limited to a certain range of crack width and crack-tip radii? If so, is it not necessary that this empirical box be closed by ascertaining the crack-tip radii limits over which the presently presented curves apply?

To approach the same question in another way, if the fatigue crack of a specified length had been produced by some other stress-cycling sequence, so as to have a different width and a different crack-tip radius, might we not expect the crack to propagate at some other maximum gross stress?

If so, the width and tip-radii limits to which the present curves apply need to be established.

Obtaining further data on the effects of altering these variables over wider ranges is, of course, a large and separate problem. The interest adhering to it does not detract in any way from the great interest inherent in the valuable data presented in the paper.

AUTHOR'S CLOSURE

The author is taking this opportunity to thank Professors McClintock and Orowan as well as Dr. Tilles for their kind discussions.

The effect of thickness, particularly with regard to plate rather than sheet thicknesses has been demonstrated by Sorensen (3) in unpublished data. Tests of thickness effect on Alclad Aluminum sheets (0.024 to 0.128 in.), with which the author is familiar, showed only slight variations in strength. These variations are probably due to the different amounts of cladding above and below 0.064-in. thick Alclad sheets.

The author appreciates Professor Orowan's amplifications on the questionable use of dW/dA and r' as material constants in crack propagation problems. For the sake of brevity the complete derivations of the equations were referenced rather than repeated. While it is agreed that a critical state of deformation in the immediate vicinity of the crack tip is a necessary condition for crack propagation, the experimental determination of such a criterion is exceedingly difficult. The size of the plastic region in steel which has been given (11) as 0.3 to 0.5 mm presents much too limited information for engineering solutions to fracture problems.

The plastic-elastic instability of a ductile fracture described (1) by the equation

$$\frac{d^2W}{dx^2} = -\frac{d^2U}{dx^2} \quad \dots \dots \dots \dots \dots \dots [14]$$

as the tangent criterion of instability was found to be unusable for experimental crack propagation studies. Despite painstaking experimental procedures including the insertion, between specimens and testing machines, of a special device for eliminating load-drop transfer from the machine to the specimens during rapid crack growth, it was found exceedingly difficult to obtain a reliable tangent criterion from the load-extension curves. If, as suggested, the ductile breaking force cannot be obtained from this criterion, then the validity of Equation [14] as a ductile fracture criterion is as yet to be demonstrated.

Dr. Tilles' inquiry regarding the importance of differences in actual and effective crack tip radii is very much to the point. The effective radius r' based on an elastic stress concentration factor becomes an artificial quantity in view of the plastic deformation in the metal surrounding the crack tip, as illustrated by the author and further explained by Professor Orowan's discussion. The difference in actual radius at the end of the saw cut (from 0.001 to 0.004 in.) does not particularly influence the stress level at initial crack growth and at the point of maximum stress is of no consequence since the crack tip is now a sharp V-notch. Any numerical similarity between r and r' is only coincidental. It has also been established that the maximum load condition is hardly affected by the process of making the initial slot. Cracks made by saw cuts, milling, or fatigue loading behaved similarly once critical crack length, i.e., instability, has been reached. As a point of interest it was also found that variations such as saw cuts with large circular and rectangular cut outs in the center yielded essentially the same test results as plain saw cuts of the same length. This serves as a further indication that the over-all crack length and state of deformation at the crack tip constitute the critical criteria for sudden fracture.

The author's experience in finding quick and almost unquestioned acceptance of these so-called material constants was one of the reasons for demonstrating, from experimental data, the large variations causing erroneous predictions and results.

Analysis of Residual Stress in Ground Surfaces of High-Temperature Alloys

By R. D. HALVERSTADT,[1] CINCINNATI, OHIO

Residual stress caused by grinding high-temperature alloys has proved to be a troublesome problem both from a standpoint of distortion of the parts and from reduced endurance limits. A complete analysis was made which measured the effect of grinding-wheel speed, grinding-wheel hardness, grinding fluid, down feed, and work speed on the residual-stress level in the surface of three alloys used in modern aircraft gas-turbine design. The results of the study show that stresses can be minimized by using lower wheel speeds and down feeds, increasing work speed and using a sulfurized oil as the grinding fluid.

Nomenclature

The following nomenclature is used in the paper:

$sfpm$ = surface feet per minute
$chem$ = chemical emulsion-water mixture
$sulf$ = sulfurated grinding oil
sol = soluble-oil-water mixture
H = thickness of test bar after grinding
Δh = thickness of any layer considered
h = distance from ground surface to any layer
f = total deflection of test bar
Δf = change in deflection
L = $^{1}/_{2}$ distance between gage points on test bar
E = modulus of elasticity
$\Delta f / \Delta h$ = slope of deflection versus depth curve
σ_n = residual stress in any layer
g = grinding ratio = $\dfrac{\text{vol work ground}}{\text{vol wheel worn}}$

Introduction

Residual-stress analysis is rapidly becoming a much-used, practical production technique rather than a method used only in the research laboratory. The widespread use of this technique has become extremely important today as the free world searches for more advanced jet-engine alloys that must operate at higher and higher temperatures. This need for residual-stress analysis has been accented by:

1 The apparent vulnerability of high-temperature alloys to residual processing stresses.

2 The need for higher design stresses in producing lighter airframes and engines.

3 The lighter cross sections of most jet-engine parts; hence stresses are more objectionable because of higher distortions.

The work described in this paper was a quantitative analysis of the residual-stress levels resulting from grinding three different alloys. It must be pointed out that the author and others in his

company have solved many problems using only a qualitative approach. These difficulties were manifested by actual cracking or checking of the ground surfaces. The surface "checking" was usually eliminated in the laboratory by adjusting the grinding variables and subsequently applying the same principles in the machine shop. This method was sufficient to produce parts which could pass a physical inspection but gave no assurance of a low level of residual stress—a quantitative approach was required.

Previous work by Tarasov and Grover (1)[2] and numerous others has shown that tensile stresses tend to reduce the endurance limit, while compressive stresses usually raise the endurance limit. In alloys with high-temperature application, the practice of shot-peening or otherwise obliterating high tensile stresses with compressive stresses is sometimes dangerous because of the possibility of recrystallization. The latter method is promoted by the added surface energy in the compressed cold-worked surface, causing recrystallization to occur at a temperature lower than normal. In severe instances of cold work, the temperature where recrystallization occurs can be very close to the operating temperature. Thus a low-strength surface would be the result, instead of the desired effect of higher strength.

The problems which prompted this study stemmed from many different types of grinding. These types included plunge grinding of complicated forms, conventional surface grinding of flat surfaces on reciprocating surface grinders, and flat grinding on rotary surface grinders which grind with the face of the wheel. It was decided to use conventional surface grinding for the tests because flat strips which are easily analyzed can be readily produced with this method.

Material Selection

The analysis was conducted on three different types of alloys used in aircraft-engine manufacture. They are:

1 Martensitic stainless, Lapelloy (G-E Spec. No. B50R311).

2 Precipitation hardening alloy, A-286 (G-E Spec. No. B50R313).

3 Precipitation hardening alloy, M-252 (G-E Spec. No. B50R327).

These three alloys were chosen because they represent typical alloys in three different operating temperature ranges.

Test Procedure

A—Material

1 A-286, $^{1}/_{4} \times {}^{3}/_{4} \times 6$-in. samples cut from $^{3}/_{4}$-in. bar stock.

2 M-252, $^{1}/_{4} \times {}^{3}/_{4} \times 6$-in. samples cut from $^{3}/_{4}$-in. bar stock.

3 Lapelloy, $^{1}/_{4} \times {}^{3}/_{4} \times 6$-in. samples cut from 1-in. bar stock.

B—Sample Preparation

1 Samples were machined and ground to 0.200 in. thick $\times$ $^{3}/_{4}$ in. wide $\times$ 6 in. long and a 10-deg angle was cut on the sides for purposes of holding samples during final grinding.

2 After machining and grinding, the samples were mechanically straightened on an arbor press before heat-treating.

3 The heat-treating was accomplished with the samples in

the "supported" position so that distortion during heat-treating would be minimized.

(a)—*A-286 Heat-Treatment.* Solution treat at 1800 F for 2 hr; water quench; age at 1300 F for 16 hr, air cool; hardness, 67 R_a average.

(b) *M-252 Heat-Treatment.* Solution treat at 1950 F for 4 hr; air cool; age at 1400 F for 15 hr, air cool; hardness, 68 R_a average.

(c) *Lapelloy Heat-Treatment.* Preheat at 1425 F for 15 min in salt bath; transfer to 2000 F for 20 min in salt bath; quench at 650–675 F for 3 min in salt bath, air cool; temper at 1330 F for 2 hr, air cool; hardness, 65 R_a average.

The intention of the foregoing heat-treatments was to produce a "stress-free" sample in the desired condition of hardness. In each instance, a thin 0.060-in-thick specimen was taken through the prescribed heat-treatment, and later investigated for stress by etching successive layers from the surface. No change in deflection from the successive etching indicated that the sample was "stress-free."

C—Grinding Procedure

All grinding was done on a 6-in. × 18-in. surface grinder equipped with an infinitely variable-speed drive. The samples were held during grinding by a special fixture as shown in Fig. 1.
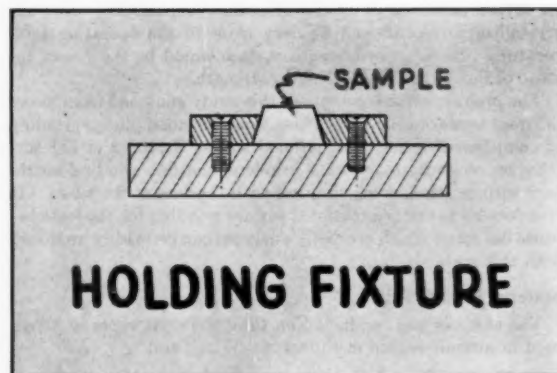
Fig. 1 Holding fixture for residual stress specimen

The 0.200-in-thick samples were mounted carefully in the test fixture and ground to 0.120-in. thickness using the grinding combination under investigation. The measurements taken before, during, and after this step, were used to calculate the grinding ratio g. At this point, the wheel was carefully redressed and an additional 0.020 in. was ground from the surface of the sample. Each test was made under equivalent amounts of stock ground from the sample. Great care was taken to insure repeatability of the dressing method. The diamond was inclined at a 20-deg angle to the face of the wheel and was continually rotated to insure that the sharpness of the point was identical in all tests. In each dress the same number of roughing and finish-dressing passes were taken at identical traverse speeds.

D—Selection of Grinding Variables

The variables investigated on each material were: Wheel speed, table speed, depth of cut, wheel-hardness grade, and grinding fluid. It should be noted that this test was not designed to explore the effects of small changes in the variables. This approach attempted only to find out what happens when large

changes in each variable were made. This form was desirable because it lent itself to analysis and comparison of results.

Blocks were set up which explored wheel speed versus type of grinding fluid, wheel speed versus hardness grade, and table speed versus down feed. The following levels of each variable were used:

Wheel speed, sfpm	2000	4000	6000
Hardness grade	G	I	K
Down feed, in.	0.0005	0.00225	0.004
Table speed sfpm	5	27	50
Fluid	Chemical emulsion with water 20:1	Soluble oil with water 20:1	Sulfurated oil

Several tests were repeated so that the effect of a short-time stress-relief heat-treatment could be ascertained.

The test conditions are shown in Table 1, 2, and 3. The tables also show the results of each test in terms of finish, g-ratio, and deflection of the 6-in. samples.

Residual-Stress Analysis

The general method of quantitative stress analysis was developed by Frisch and Thomsen (2). The residual stress imparted by the grinding causes the strip to bend and the total deflection between two gage points can be measured. The deflection was measured by a gage as shown schematically in Fig. 2. When
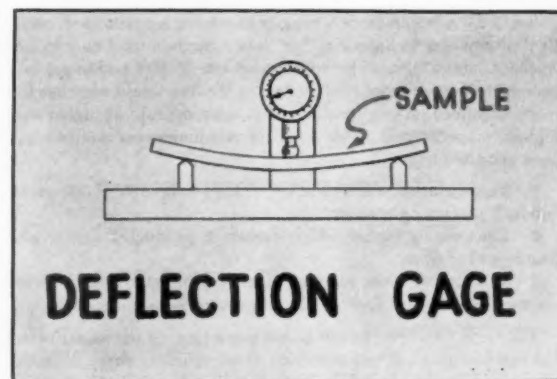
Fig. 2 Deflection gage for measuring change in deflection

thin layers of stressed material are etched from the surface of the bent strip, the strip changes deflection which was measured by the gage. The layer removal was continued until no further deflection change was apparent. These data were then used to plot the change in deflection of the strip as a function of the depth below the surface. The slopes from this curve are then used to calculate the stress, utilizing the Frisch and Thomsen equation (2)

Table 1 Lapelloy

(Grinding wheel: friable, aluminum oxide, 60 grit, 6 spacing, vitrified bond. Except as noted: work speed: 15 sfpm; down feed: 0.002 in.; cross feed: 0.030 in.; coolant: chemical emulsion-water, 20:1.)

Wheel "I"

Coolant		Wheel speed, sfpm		
		2000	4000	6000
Sulfurated oil	Bar No.	L-5		L-4
	Deflection	+0.0012		+0.0182
	Finish	90		20
	g-ratio	177		118
Sol. oil 20:1	Bar No.		L-1	
	Deflection		+0.0034	
	Finish		23	
	g-ratio		32.0	
Chemical emulsion-water 20:1	Bar No.	L-3	L-10	L-2, L-16,[a] L-17[a]
	Deflection	+0.0015	+0.0028	+0.0175 ⎫
	Finish	70	22	15 ⎬ for L-2
	g-ratio	8.68	23.8	36.0 ⎭

Wheel hardness		Wheel speed, sfpm		
		2000	4000	6000
G	Bar No.	L-6		L-9
	Deflection	+0.0004		+0.0075
	Finish	100		20
	g-ratio	1.50		10.1
I	Bar No.	L-3	L-10	L-2
	Deflection	+0.0015	+0.0028	+0.0175
	Finish	70	22	15
	g-ratio	8.68	23.8	36.0
K	Bar No.	L-7		L-8
	Deflection	+0.001		+0.0151
	Finish	90		22
	g-ratio	0.55		34.6

Wheel "G," speed 4000 sfpm

Down feed, in.		Work speed, sfpm		
		5	27	50
0.0005	Bar No.	L-11		L-14
	Deflection	+0.001		+0.0013
	Finish	9.5		35
	g-ratio	46.9		32.8
0.00225	Bar No.		L-13	
	Deflection		+0.0016	
	Finish		70	
	g-ratio		8.50	
0.004	Bar No.	L-12		L-15
	Deflection	+0.0046		+0.0029
	Finish	15		120
	g-ratio	53.3		0.590

NOTE: Positive deflection indicates an over-all tensile effect; finish was measured in microinches (arithmetic mean average).

Deflection measured in inches; distance between gage points, $5\frac{1}{2}$ in.

[a] Bar L-16 and L-17 used for stress relief test.

$$\sigma_n = \frac{E}{L^2}\left[\frac{(H - h_n)^2}{3\left(1 + \frac{\Delta h_n}{H - h_n}\right)}\left(\frac{\Delta f}{\Delta h}\right)_n - (H - 2h_n + \Delta h_n)f_n - 1\right.$$

$$\left. - \sum_0^{n-1}\left\{\frac{(H - h)^2}{3\left(1 + \frac{h}{H - h}\right)}\left(\frac{\Delta f}{\Delta h}\right)\left(\frac{\Delta h}{H - h}\right) + h\Delta f\right\}\right] \dots [1]$$

A complete description of the method can be obtained from reference (2). The equation uses h in place of Δh in the denominator of the term $\Delta h/(H - \Delta h)$ in the preceding summation. It is believed that this was an error in the original derivation. The equation was programmed on a computer and, thus, the calculations could be made quite rapidly. Many extra computations in areas of sharp stress transition were possible which could not have been made without this labor-saving device. For complete deriva-

tion of Equation [1] the reader should refer to (2) in the Bibliography.

Etching Procedure

The layer removal was accomplished on all three materials using an electrolytic swab method with 15–20 per cent hydrochloric acid as the electrolyte at six volts.

The work of etching and measuring deflection was done by Dr. Michael Field and his associates at Metcut Research Associates Incorporated.

The samples were measured accurately before and after etching so that the measurement of the thickness removed was representative. Measurements were actually taken at 21 different positions on the surface, and if any spot measured "heavy," the electrolytic swab was used to remove the material on the localized area.

Table 2   A-286

(Grinding wheel: friable, aluminum oxide, 60 grit, 6 spacing, vitrified bond. Except as noted: work speed: 15 sfpm; down feed: 0.002 in.; cross feed: 0.030 in.; coolant: sulfurated oil.)

Wheel "I"

Coolant		Wheel speed, sfpm		
		2000	4000	6000
Sulfurated oil	Bar No.	A-10	A-15	A-9, A-23,[a] A-25,[a] A-27[a]
	Deflection	+0.0081	+0.0247	+0.0635
	Finish	50	24	20 } for A-9
	g-ratio	11.3	11.5	36.8
Sol. oil 20:1	Bar No.		A-3	
	Deflection		+0.0486	
	Finish		45	
	g-ratio		20.3	
Chemical emulsion-water 20:1	Bar No.	A-7	A-8	A-6
	Deflection	+0.0174	+0.0435	+0.0447
	Finish	40	28	25
	g-ratio	0.58	2.43	4.61

Wheel hardness		Wheel speed, sfpm		
		2000	4000	6000
G	Bar No.	A-11		A-14
	Deflection	+0.0074		+0.0372
	Finish	45		17
	g-ratio	5.4		53.0
I	Bar No.	A-10	A-15	A-9
	Deflection	+0.0081	+0.0247	+0.0635
	Finish	50	25	20
	g-ratio	11.3	11.5	36.8
K	Bar No.	A-12		A-13
	Deflection	+0.0084		+0.0581
	Finish	35		12
	g-ratio	10.4		74.0

Wheel "G"

Down feed, in.		Work speed, sfpm		
		5	27	55
0.0005	Bar. No.	A-16		A-19
	Deflection	+0.0087		+0.0057
	Finish	13		15
	g-ratio	13.1		10.0
0.00225	Bar No.		A-18	
	Deflection		+0.0099	
	Finish		37	
	g-ratio		6.9	
0.004	Bar No.	A-17		A-20
	Deflection	+0.0359		+0.0113
	Finish	20		65
	g-ratio	24.6		0.55

NOTE: Positive deflection indicates an over-all tensile effect; finish was measured in microinches (arithmetic mean average).

Deflection measured in inches; distance between gage points, 5¹/₂ in.

[a] A-23, 25, 27 used for stress relief and crosswise grinding tests.

Metallurgical Studies

An effort was made to determine the depth of "cold work" in the ground specimens by conducting recrystallization tests. An attempt was made by Metcut Research to develop a recrystallization test for Lapelloy, but this was not feasible. The tests on A-286 and M-252 were conducted at 1500 and 1600 F for 1 hr, respectively.

Results

The tabulated results in terms of finish, grinding ratio "g," and deflection on a 6-in. sample, are shown in Tables 1, 2, and 3. The results of the quantitative residual-stress analysis are shown in Figs. 3 through 30.

It is important to note that the analysis was not made on every ground sample shown in the Tables 1, 2, and 3. Representative samples are shown and the deflection of the samples can be correlated with the stress analysis to synthesize the stress pattern on the remaining samples.

The results of the stress analysis are presented by a series of comparisons, and the *differences in grinding conditions on the samples presented in any one graph are noted*. For a complete history of the conditions it is necessary to refer to Tables 1, 2, and 3.

The results for each material will be presented separately, and a final comparison will be made which shows the difference between the three materials ground under equivalent conditions.

Lapelloy

Fig. 3 shows slight difference between the chemical emulsion-water mixture and the sulfurated grinding oil when the samples

Table 3   M-225

(Grinding wheel: friable, aluminum oxide, 60 grit, 6 spacing, vitrified bond.   Except as noted: work speed: 15 sfpm; down feed: 0.002 in.; cross feed: 0.030 in.; coolant: sulfurated oil.)

Wheel "I"

Coolant		Wheel speed, sfpm		
		2000	4000	6000
Sulfurated oil	Bar No.	M-5	M-6	M-4, M-16,[a] M-17[a]
	Deflection	+0.0050	+0.0109	+0.0451
	Finish	37	25	20
	g-ratio	21.4	9.4	64.5
Sol. oil 20:1	Bar No.		M-1	
	Deflection		+0.0208	
	Finish		45	
	g-ratio		2.6	
Chemical emulsion-water 20:1	Bar No.	M-3		M-2
	Deflection	+0.0140		+0.0441
	Finish	37		32
	g-ratio	2.1		4.73

Wheel hardness		Wheel speed, sfpm		
		2000	4000	6000
G	Bar No.	M-8		M-7
	Deflection	+0.0018		+0.0121
	Finish	40		24
	g-ratio	8.4		61.3
I	Bar No.	M-5	M-6	M-4
	Deflection	+0.0050	+0.0109	+0.0451
	Finish	37	25	20
	g-ratio	21.4	9.4	64.5
K	Bar No.	M-9		M-10, M-20
	Deflection	+0.0036		+0.0423 } for M-10
	Finish	40		15
	g-ratio	18.4		45.7

Wheel "G," speed 4000 sfpm

Down feed, in.		Work speed, sfpm		
		5	27	50
0.0005	Bar No.	M-15		M-12
	Deflection	+0.0024		+0.0025
	Finish	30		17
	g-ratio	58.0		13.8
0.00225	Bar No.		M-13	
	Deflection		+0.0091	
	Finish		28	
	g-ratio		11.9	
0.004	Bar No.	M-14		M-11
	Deflection	+0.0249		+0.0159
	Finish	20		70
	g-ratio	77.0		0.57

NOTE: Positive deflection indicates an over-all tensile effect; finish was measured in microinches (arithmetic mean average).

Deflection measured in inches; distance between gage points, $5\frac{1}{2}$ in.

[a] Bar M-16 and M-17 used for stress relief tests.
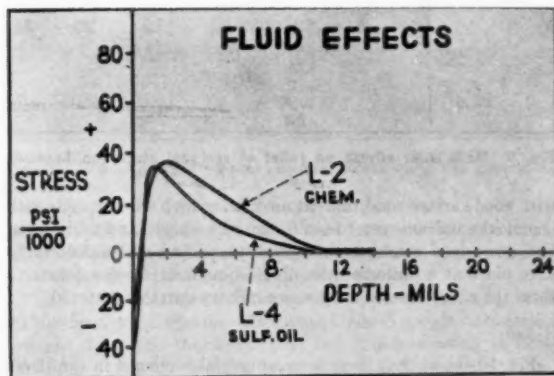
Fig. 3   Grinding-fluid effects on residual stress on Lapelloy
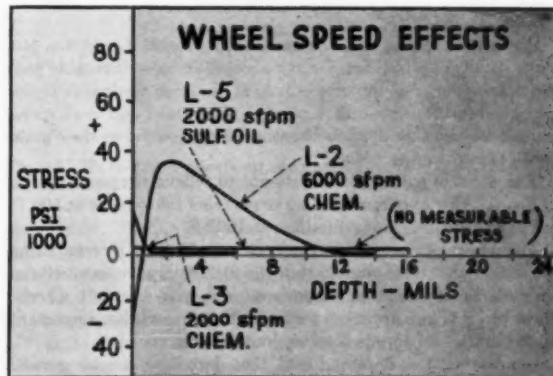
Fig. 4   Grinding-wheel speed effects on residual stress on Lapelloy

Fig. 5 Work-speed effects on residual stress on Lapelloy ground at low down feeds
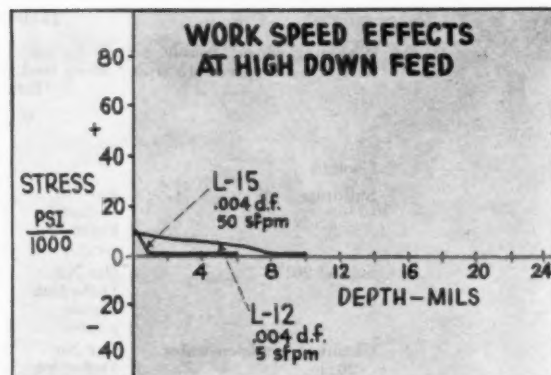
Fig. 6 Work-speed effects on residual stress on Lapelloy ground at high down feeds

Fig. 7 Down-feed effects on residual stress on Lapelloy ground at low work speeds
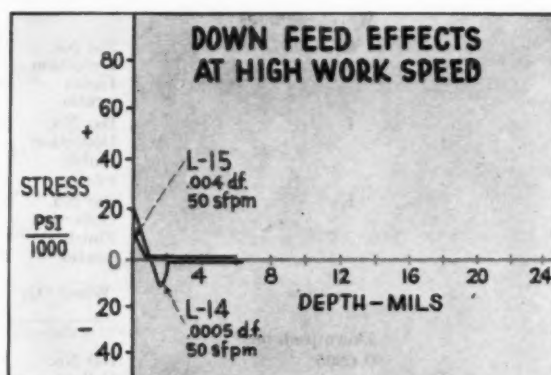
Fig. 8 Down-feed effects on residual stress on Lapelloy ground at high work speeds

were ground at 6000 sfpm, but Fig. 4 shows that the sulfurated oil is superior at the low wheel speed.

Fig. 4 shows the effect of changing wheel speed from 6000 to 2000 sfpm; both the level and depth of stressed area are materially reduced.

The effect of work speed can be seen in Figs. 5 and 6. The effect is not clean-cut, but the high work speed produced the most favorable results at the heavy down feed. The opposite is true at the low down feed. There is evidently a balance between down feed and work speed that will produce the optimum condition of chip size and grinding conditions to minimize residual stress.

Down-feed effects can be seen in Figs. 7 and 8. At the low work speed the low down feed produces the most favorable pattern; however, the opposite is true at the high work speed. As suggested in the preceding paragraph, down feed and work speed cannot be separated since together they determine the "grain depth of cut" and chip size.

The effect of a 1-hr heat-treatment to relieve stresses is shown in Fig. 9. It is apparent that no stress relief takes place at 500 F, but considerable relief is obtained at 1000 F.

The wheel-hardness effect can be seen in Table 1 by comparing the deflections. It is obvious that the surface stresses as measured by deflections are lower when produced with the soft wheels. However, it is also apparent that wheel hardness is less important at the low wheel speeds than at conventional speeds.

In general, the data indicate that Lapelloy can be ground
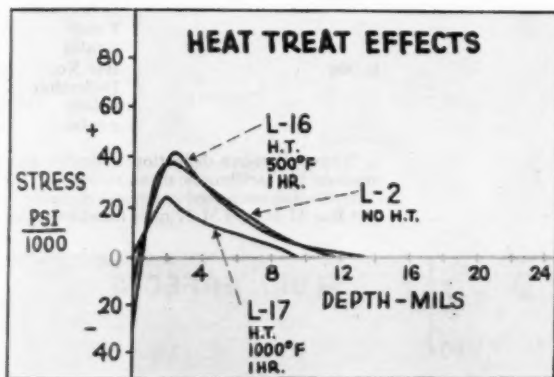
Fig. 9 Heat-treat effects on relief of residual stress on Lapelloy

with good surface conditions at medium to high wheel speeds with chemical emulsion-water base fluids, soft wheels, and table speed and down-feed combinations which permit wheel breakdown to take place at a uniform rate. Not enough data were taken to show the effect of soluble oil-water mixtures on this material.

A-286

Fig. 10 shows that there is an appreciable change in the stress
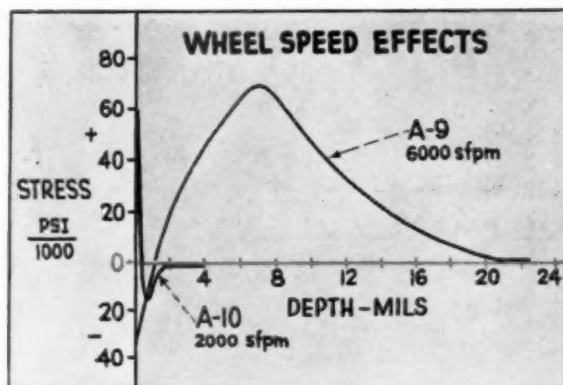
Fig. 10 Grinding wheel speed effects on residul stress on A-286

Fig. 11 Grinding-fluid effects on residual stress on A-286 ground at high wheel speeds

Fig. 12 Grinding-fluid effects on residual stress on A-286 ground at low wheel speeds

Fig. 13 Grinding-fluid effects on residual stress on A-286 ground at intermediate wheel speeds.
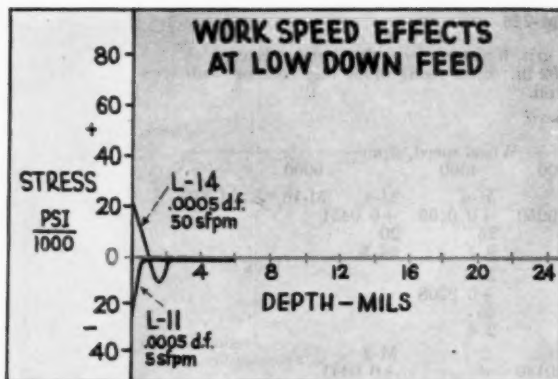

Fig. 14 Work-speed effects on residual stress on A-286 ground at low down feeds.

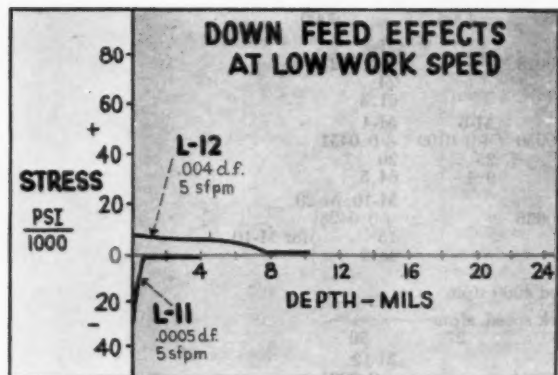Fig. 15 Work-speed effects on residual stress on A-286 ground at high down feeds

pattern when the wheel speed is reduced from 6000 to 2000 sfpm. With the low speed there is a rather high residual stress present, but the depth of penetration is greatly reduced.

Figs. 11, 12, and 13 show the results of the different fluids at the different wheel speeds. At the high wheel speeds both stress patterns show high tensile stresses, and it is interesting to note the absence of a compressive layer with the chemical emulsion-

water fluid. At the low wheel speed the lowest stress was produced with the sulfurated oil. A comparison of the chemical emulsion-water base fluid and soluble oil-water base is shown in Fig. 13. Soluble oil appears to be the least desirable of the two grinding fluids.

Figs. 14, 15, 16, and 17 show the effects of various combinations of down feed and work speed. It is apparent that if a low down
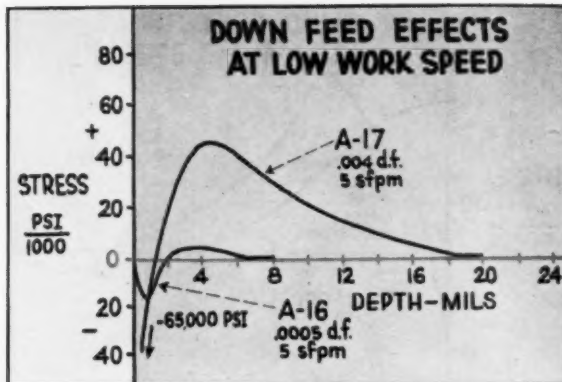
Fig. 16 Down-feed effects on residual stress on A-286 ground at low work speeds
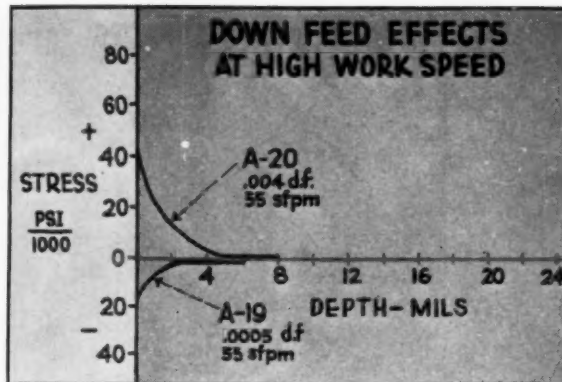
Fig. 17 Down-feed effects on residual stress on A-286 ground at high work speeds
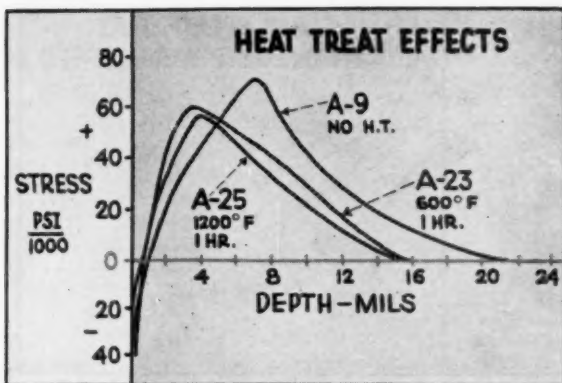
Fig. 18 Heat-treat effects on relief of residual stress on A-286
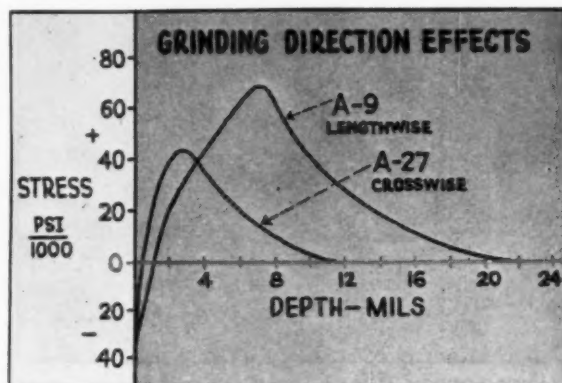
Fig. 19 Grinding-direction effects on residual stress on A-286

feed is used, the work speed can be varied considerably and still produce a desirable surface condition. However, at the high down feeds, a high work speed is mandatory for improved surface stress conditions. It is obvious that the down feed has a pronounced effect as shown in Figs. 16 and 17.

The 1-hr stress-relief heat-treatment results are shown in Fig. 18. Neither specimen exhibited any appreciable amount of stress relief, although sample A-25 is slightly lower. The latter effect is within the probable experimental repeatability of the tests, and therefore it is concluded that slight relief was attained.

The results of grinding one sample "crosswise" are shown on Fig. 19. A stress pattern similar to the lengthwise pattern was obtained, but since the wheel-breakdown conditions are completely different, the same stress level was not expected.

In general, lower wheel speed and down feed and increased work speeds used with a sulfurated-oil fluid are indicated. Wheel hardness, as shown in Table 2, appears to have little effect at 2000 sfpm, but soft wheels produced the best results at 6000 sfpm.

M-252

The wheel-speed effect is more pronounced on M-252, as shown in Fig. 20. The complete absence of a measurable stress on sample M-5 indicates that speed change is a powerful tool in the elimination of residual grinding stresses.

The comparison of sulfurated grinding oil with the chemical emulsion-water fluid at two different speeds is shown in Figs. 21 and 22. It is apparent that the oil is superior at the lower speed.
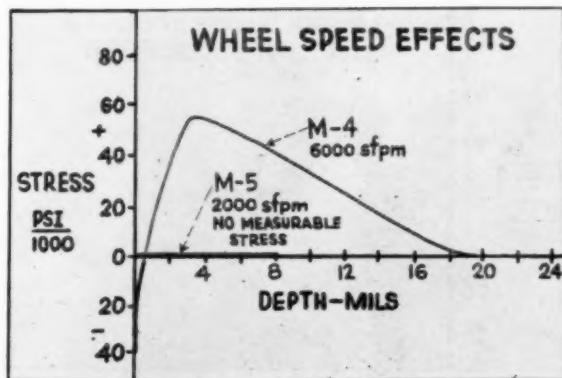
Fig. 20 Grinding-wheel speed effects on residual stress on M-252

The various combinations of work speed and down feed are shown in Figs. 23, 24, 25, and 26. With the low down feed the extremes in work speed had little effect. However, the use of high work speed appears mandatory if a "heavy" down feed is utilized. As in A-286, the down feed appears to be a powerful tool in reducing stresses.

The samples which were heat-treated are shown in Fig. 27
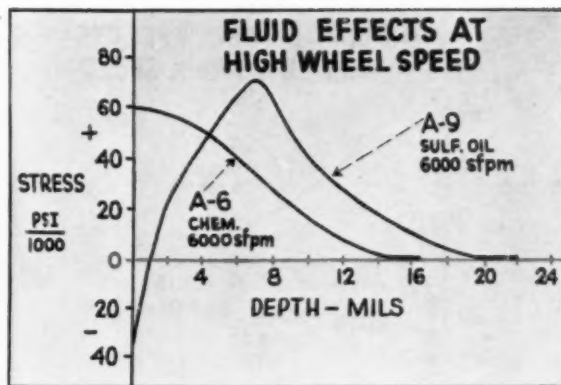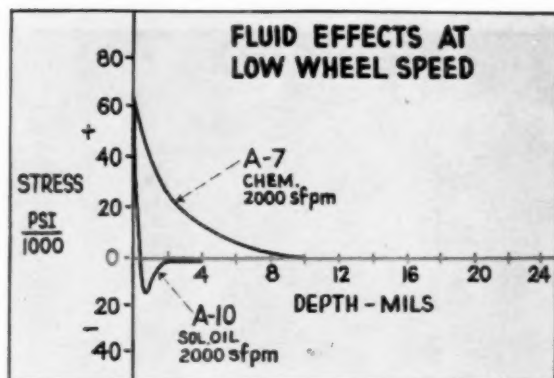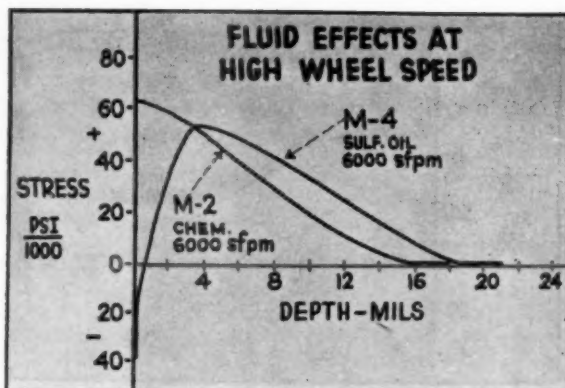
Fig. 21   Grinding-fluid effects on residual stress on M-252 ground at high wheel speeds
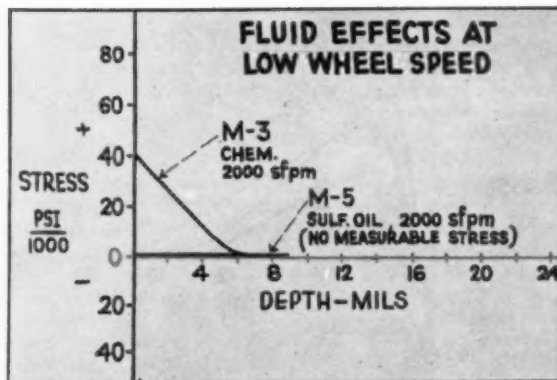
Fig. 22   Grinding-fluid effects on residual stress on M-252 ground at low wheel speeds
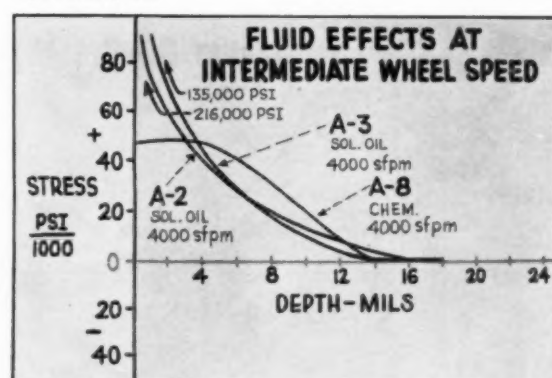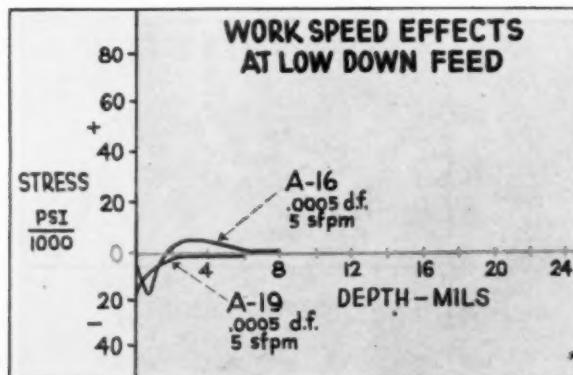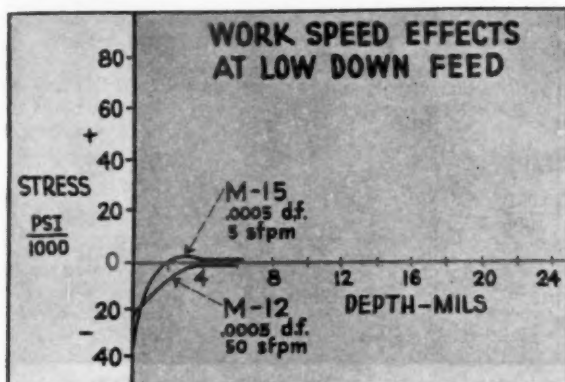

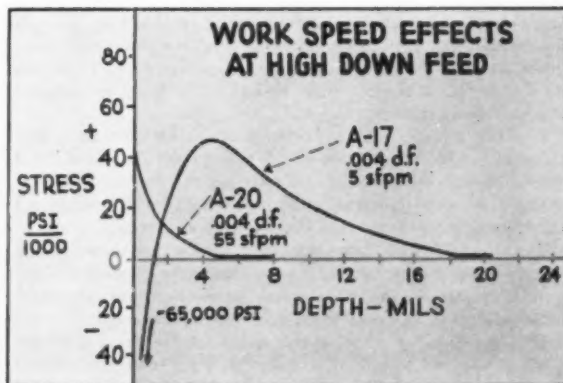Fig. 23   Work-speed effects on residual stress on M-252 ground at low down feeds
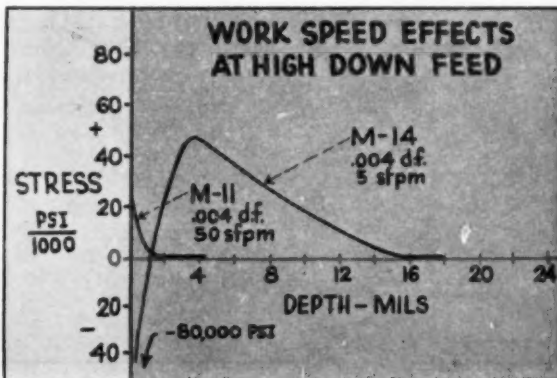
Fig. 24   Work-speed effects on residual stress on M-252 ground at high down feeds
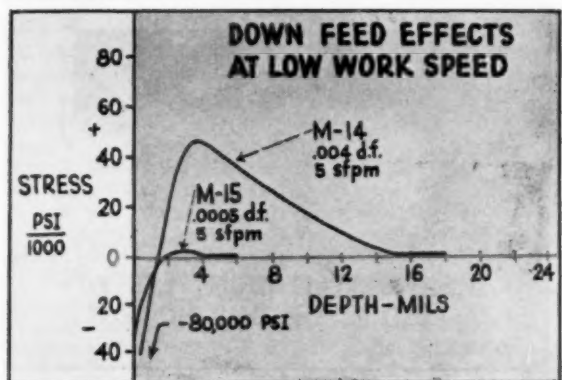
Fig. 25   Down-feed effects on residual stress on M-252 ground at low work speed
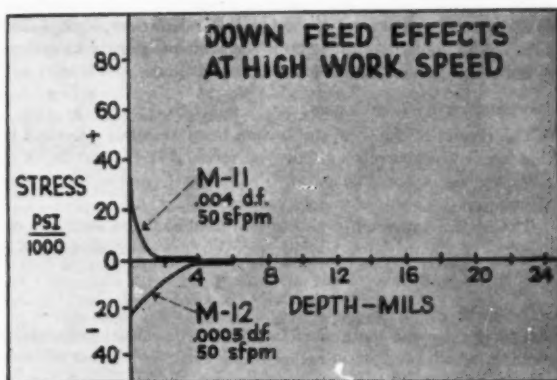
Fig. 26   Down-feed effects on residual stress on M-252 ground at high work speeds

There is no relief with the 700 F treatment, but considerable relief was obtained on the sample treated at 1400 F.

As with A-286, the "crosswise" ground sample, shown in Fig. 28, again exhibited the same general pattern as the longitudinal ground bars; apparently the difference in grinding conditions caused a reduction in stress level, although Letner (3, 4) has shown that a lower level would be expected.

As in the two other alloys, Table 3 indicates that wheel hardness is important at the high wheel speeds, where the "G" wheel resulted in the lowest deflection, but is not as important at the low wheel speeds.

Comparison of Materials

Figs. 29 and 30 show a comparison of all three materials ground
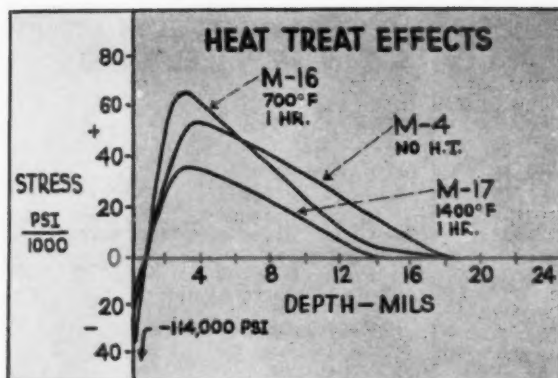
Fig. 27   Heat-treat effects on relief of residual stress on M-252
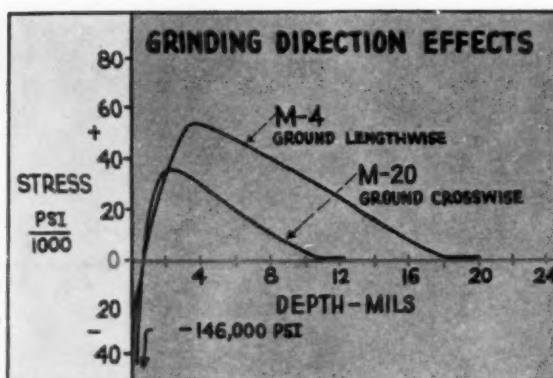
Fig. 28   Grinding-direction effects on residual stress on M-252
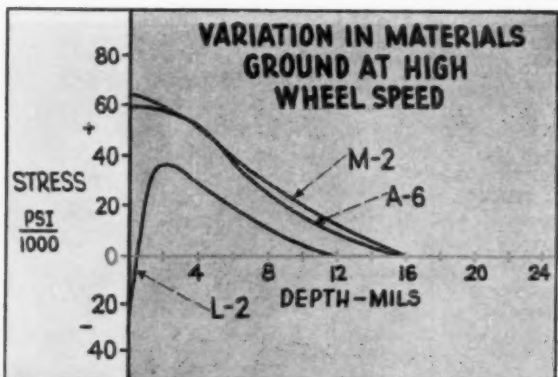
Fig. 29   Variation of residual stress on the three materials ground at high wheel speeds
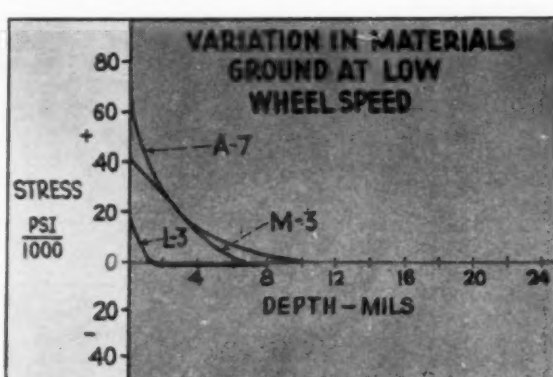
Fig. 30   Variation of residual stress on the three materials ground at low wheel speeds

under the same conditions. Lapelloy, is certainly less vulnerable to stresses than either of the other two. However, both A-286 and M-252 appear to have about the same propensity to surface stress when ground under identical conditions.

Recrystallization Test Results

The results of the recrystallization tests were not revealing in that no distinct correlation between residual stress and depth of recrystallization could be made. The A-286 showed no correlation although it did occur at widely varying depths.

The M-252 appeared to be recrystallized at the same depth regardless of the grinding conditions. A representative sample is shown in Fig. 31.

Discussion

The electroswab method of layer removal caused preferential etching, which resulted in rather rough surfaces. These surfaces were somewhat difficult to measure accurately. However, the deflection measurements were all taken from the smooth side. It is probable that electropolishing for removal of the layers and weighing the samples instead of measuring the thickness would have resulted in greater accuracy. It is difficult to determine positively the probable inherent error in the method, since it depends not only on the accuracy of the thickness and deflection measurements, but also on the accuracy of determining the slope of the deflection versus depth curve. It is probable that the method is accurate to within ±5000 psi, but the error could easily exceed 10,000 psi. The determination of stress at the extreme surface is subject to
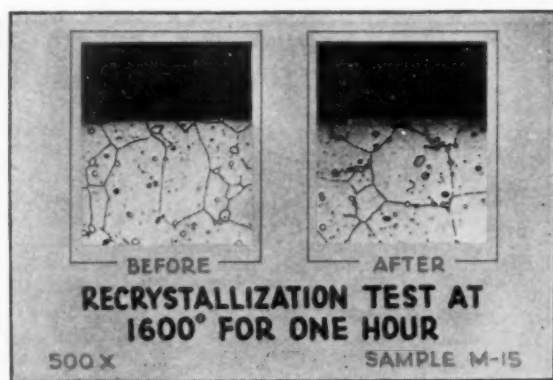
Fig. 31   Recrystallization test at 1600 deg for 1 hr on M-252 showing microstructure before and after treatment

much greater error. The latter is especially true if the stress is rapidly changing from compression to tension, or vice versa. If on the first layer removal the compressive layer is entirely removed, it is possible that its presence will not be detected at all. In the calculation of the surface stress at depth equal zero, the stress depends only on an accurate determination of the slope since the equation at depth zero simplifies to

$$\sigma_n = \frac{E}{3}\left(\frac{H}{L}\right)^2\left(\frac{df}{dh}\right)\dots\dots\dots\dots [2]$$

The slope is difficult to ascertain accurately, and therefore the surface-stress calculation tends to be somewhat more in error than the succeeding values.

The assumption of a simple uniaxial stress may also be somewhat in error, but since all measurements were taken under identical conditions and used in a comparative manner, errors arising from this cause are not objectionable.

In any event, the method is suitable for the purpose intended. It was not the intention to obtain completely precise data in absolute terms, but rather to develop figures which show the overall trends. Succeeding tests of a more detailed nature could cover the minute details and seek precision answers from which separate effects of small changes could be made.

It may be surprising to find stresses as high as those shown on samples A-2 and A-3, in Fig. 13, since these stresses are close to or exceed the room-temperature yield stress for this material. Theoretically, the residual stress cannot exceed the yield stress, since any tendency to cause stresses higher than yield will merely cause plastic flow until the yield stress is reached. The phenomena of stresses higher than the bulk yield have occurred often and have been discussed in detail by Shaw (6) and by Backer, Marshall, and Shaw (5). The stresses on these two samples are higher than those reported by Letner (3, 4) on hardened steel.

It is interesting to note that based on deflection alone, as shown in Tables 1, 2, and 3, all three materials reacted about the same to wheel-hardness grade. The "G" wheel gave the best results, but the two harder wheels showed little difference. It is unfortunate that residual-stress analysis which could be directly compared were not made for wheel hardness; however, the tests were limited by budget and time requirements.

The fact that compressive stresses only were present on bars A-19, M-15, M-12, and L-11 which corroborates the findings of Letner (3) are both interesting and useful information. This shows definitely that grinding need not be a damaging process.

The effect of grinding fluid was not particularly pronounced at high wheel speeds but was more evident at lower speeds. This could indicate that more time was available for the reaction of the various agents in the oil at lower speeds.

Wheel speed was demonstrated to be a powerful tool in reducing the stress. Some of the tests indicate that wheel speed alone practically can eliminate the stress. However, wheel speed is just one of the variables, and changes just as great can be made by other variables. As shown in the data, a good stress pattern was obtained at 4000 sfpm in all alloys. It is probable that the same could be true for 6000 sfpm if all other conditions are "ideal." The likelihood that all other conditions would often be ideal is small; therefore somewhat lower than normal wheel speeds for these alloys is recommended. Production experience has shown that satisfactory conditions have been obtained between 3000–4000 sfpm on most jobs.

The combination of work speed and down feed is of paramount importance. In general, it appears safe to assume that high work speeds should be used for finishing in conjunction with low down feeds. The low down feed in all materials appeared to be one of the surest methods of reducing the tensile-stress level. Usually, low down feeds helped promote good finish and wheel life. The grinding process is essentially a finishing process, and the necessity for low down feeds should not cause consternation. If, however, low down feeds are not available, it has been possible to reduce stress in production parts by a change in wheel speed and grinding fluid alone.

The findings derived at here have been applied to several different production grinding jobs and have invariably been success-

ful in effecting a solution to the grinding problem. In one instance, the problem was that of lowered endurance limits, which was apparently cured by the improved grinding methods. In several other instances surface "checking" was stopped by utilizing the same principles. The distortion of ground parts also proved to be troublesome, but application of these principles has eliminated the difficulty.

In general, results such as these should be used in the qualitative sense and not applied directly in the machine shop. Thus if a low wheel speed shows a major improvement, it should show improvement in the shop, although not necessarily to the same degree. The same is true of all other variables. A thorough study of the results, and a solution effected from a knowledge of the "cause-and-effect relationships," will usually be successful. A general understanding of other research work in the same field can also help to clarify the result.

Conclusions

1 The depth and level of residual stresses resulting from surface grinding the high-temperature alloys appear to be greater than for alloy steels ground under equivalent conditions.

2 Tensile residual stresses can be lowered and depth of penetration can be reduced by:

(a) Reducing wheel speed.
(b) Reducing down feed.
(c) Use of sulfurated grinding oil at optimum conditions.
(d) Use of "soft" grinding wheels.

3 Increasing work speed reduced depth of penetration in this study.

4 It is possible to grind alloys of this type with compressive stresses only.

5 Short-time heat-treatment for stress relief does not relieve all of the residual stress.

6 Addititional work should be performed on a representative alloy to determine more closely the optimum combination of variables and to explore the effects of other types of grinding.

Acknowledgment

Bibliography

1 "Effects of Grinding and Other Finishing Processes on the Fatigue Strength of Hardened Steel," by L. P. Tarasov and H. J. Grover, Proceedings of the American Society for Testing Materials, vol. 50, 1950, pp. 668–698.
2 "Residual Grinding Stresses in Mild Steel," by J. Frisch and E. G. Thomsen, Trans. ASME, vol. 73, 1951, pp. 337–346.
3 "Residual Grinding Stresses in Hardened Steel," by H. R. Letner, Trans ASME, vol. 77, 1955, pp. 1089–1098.
4 "Influence of Grinding Fluids Upon Residual Stresses in Hardened Steels," by H. R. Letner, Trans. ASME, vol. 79, 1957, pp. 149–153.
5 "The Size Effect in Metal Cutting," by W. R. Backer, E. R. Marshall, and M. C. Shaw, Trans. ASME, vol. 74, 1952, pp. 61–72.
6 "Plastic Flow in the Cutting and Grinding of Metals," by M. C. Shaw, Proceedings of the National Academy of Sciences, vol. 40, 1954, pp. 394–400.

Discussion

L. P. Tarasov.[3] The inherent complexity of grinding is clearly revealed in this study of the effect of six major variables

[3] Research and Development Department, Norton Company, Worcester, Mass. Mem. ASME.

upon residual stresses, grinding ratio, and surface roughness. With a few exceptions, the results follow a general pattern which is consistent enough to permit definite conclusions to be drawn as to how to adjust the grinding conditions to minimize the residual tensile stresses. However, they also show that the effect of a given change in grinding conditions upon the grinding ratio can be markedly different for one alloy than for another.

Most of these variations in the grinding ratio with grinding conditions appear to be consistent with one another, but there are a few which seem to be out of line. The grinding ratio for L-5 might be expected to be much lower in view of the very rough finish reported since the other cases of poor finish are associated with low grinding ratios. On the other hand, the grinding ratio appears to be too low for L-7 (K-grade) by comparison with L-3 (I-grade), inasmuch as the grinding ratios were practically the same for these two wheels when either of the other two alloys was ground at 2000 sfpm. Similar intercomparisons indicate that the grinding ratios may be too high for A-3 and M-15, and too low for A-15 and M-6. It would be desirable to obtain additional grinding-ratio data both for these conditions and for intermediate values of the appropriate variables to determine whether the apparently inconsistent values are reproducible. If they are, it would mean that the grinding action changes very rapidly but consistently within certain ranges of grinding conditions; if not, then the grinding action under these conditions must be inherently erratic, and the same is likely to be true of the residual stresses.

The results indicate that Lapelloy can be ground successfully with the particular chemical emulsion that was used, but it does not necessarily follow that this can be done with other products of this type since they include a wide variety of chemical ingredients, some of which may not be at all suitable. It would be well to state at least the general nature of the product that was used. The same applies to the soluble oil and the grinding oil.

Some additional information would be of value to the reader. This includes the nominal compositions and the yield strengths of the three alloys, and an indication of the thicknesses of the layers etched away in determining the residual stresses. It also would be helpful to state the wheel size and specifications to make the record more complete. The type of abrasive and the grit size are clear enough, but the wheel grade and spacing are somewhat ambiguous because they vary to a certain extent among wheel manufacturers. In the writer's opinion, it is always desirable to state the actual wheel markings since these provide a more specific picture of the grinding conditions.

A knowledge of the residual stresses is useful in itself since they can cause distortion and cracking. They also can affect the fatigue properties, but very little is actually known about the extent to which a given residual-stress distribution may be beneficial or detrimental from the standpoint of fatigue properties. Such a study of one or more of the high-temperature alloys would be extremely valuable since it would attach additional practical significance to the residual-stress curves already obtained. In this connection, the cold work introduced by grinding may have its own effect upon the fatigue properties of these alloys, separate from the effect of residual stresses, if the results recently obtained for hardened steel[4] are applicable to high-temperature alloys. The extent of cold work probably would have to be measured in terms of the shape of x-ray dif-

fraction lines since the recrystallization technique does not appear to be sufficiently sensitive.

Author's Closure

Dr. Tarasov has pointed out certain inconsistencies in the data and his comments are certainly valid and well taken. It is hoped that research in grinding can continue to progress to the point where it is an exact science rather than the partial "art" it is today. There has not yet been an opportunity to check the data in question but the author is attempting to refine the testing techniques so that the cause of such erratic data can be clearly determined in future studies. Many studies of the grinding process have shown that good finish and high grinding ratios certainly do appear to be directly correlated.

However, it would not be a completely safe assumption that residual stresses would also be low where good finish and high grinding ratios are found. The levels of residual stress appear to be related to the sharpness of the grinding wheel, the amount of lubrication at the grit-chip interface, and other factors such as wheel speed and down feed. In short, the grinding temperatures apparently set the level of residual stress. Thus, it has been found that residual stresses are lower for slow wheel speeds than for high wheel speeds even though both finish and g ratio are improved at the high speeds. This effect can be seen in the data presented in this paper in Tables 1, 2, and 3.

The omission of certain data was partially intentional because of proprietary reasons and in others purely unintentional. The data which can be supplied is given below:

Wheel size: 7 in. diam by $1/2$ in. wide

Additional nominal properties of the three alloys are:

	Lapelloy	A-286	M-252
Yield strength (0.02% offset at 100 F)	110,000 psi	87,000 psi	102,000 psi
Mod. of elasticity	30×10^6 psi	28.7×10^6 psi	29.95×10^6 psi
Nominal compositions, %	C, 0.30	C, 0.08	C, 0.12
	Cr, 12	Mn, 1.5	Cr, 19
	Ni, 0.50	Si, 0.7	Ni, 53
	V, 0.25	Cr, 15.0	Mo, 10
	Mo, 2.8	Ni, 25.0	Co, 10
	Mn, 1.0	Mo, 1.25	Ti, 2.5
	Si, 0.40	Ti, 2.0	Al, 1.0
	Fe, Bal.	Al, 0.35	Fe, 3.0
		V, 0.25	
		Fe, Bal.	

The layers removed during the residual stress analysis varied as the total amount of stock removed increased. The first layers removed were from 0.0001 in. to 0.0002 in. in thickness for the first 0.001 and then the thickness increments were increased to around 0.0005 up to the depth where no further deflection change was noted.

Dr. Tarasov suggests that a study to determine the effects of the various residual stresses on endurance limits should be conducted. The author is now co-operating with others in his company making just such a study and it is hoped that the results can be reported at some future time.

---

[4] "Effect of Grinding Conditions and Resultant Residual Stresses on the Fatigue Strength of Hardened Steel," by L. P. Tarasov, W. S. Hyler, and H. R. Letner, 1957 preprint, American Society for Testing Materials.

General Design Considerations for Smaller Gas Turbines

By W. T. VON DER NUELL,[1] LOS ANGELES, CALIF.

Following the general acceptance of its big brother, the smaller gas turbine is entering the market as a strong contender among power producers. While many thousands of units in the range up to 500 hp are in daily use, especially in the Armed Forces, new developments are active on a large scale with the biggest financial input probably aiming at vehicle propulsion. Due to the flexibility in arrangement and working the smaller gas turbine can find application in numerous fields where its outstanding capabilities offer definite advantages. Recent developments have demonstrated that its fuel consumption need not be too high and that it can be a very reliable machine. For further improvements, the most important contributions can be expected through better heat exchangers and better materials in addition to the continued efforts to raise component performance and to develop greater adaptability, such as reversibility.

Introduction

WHEN excluding the millions of automobiles, the largest amount of power produced comes from the shafts of turbines supplied with water, steam, or gas. Yet, only about half a century has passed since this type of mankind's servant began to attract industrial attention. The accelerated introduction of the gas turbine as an aircraft prime mover, dictated by war competition and supported by almost unlimited funds, contributed greatly to the extension of its application to many other fields. For years probably, almost all technical institutions have been working on subjects relating to gas turbines in one form or another. Each year, power plants whose turbine shafts transmit upward of 5000 hp are being manufactured by the tens of thousands; in the air forces of a single nation, there are gas turbines with a total of over 25×10^6 hp.

One of the advantages of turbines is their capability of consuming the prepared working substance at great through-flow velocities with pleasant operational smoothness. Having become a standard for stationary installation, they are gaining ever-increasing interest for mobile equipment with economical considerations of all-embracing scope determining success or failure believed to be in their favor. The two outstanding aspects are technical promise and/or perfection and manufacturing feasibility.

Very good over-all performance characteristics have been demonstrated and improvements are progressing well, enhanced by advanced knowledge in the field of gas dynamics. As for manufacturing, at the risk of slight exaggeration, can it not be said that production engineers always have and always will find methods of manufacturing products economically when there exists a genuine demand and the required materials are available?

Now then, the link connecting scientific knowledge and pro-

duction know-how appears to be the design, i.e., the translation of ideas into objects which can be manufactured profitably and will do a better job. Therefore, let us consider our subject, the smaller gas turbine, more from a designer's viewpoint.

Having been active in turbomachinery and related engineering, particularly turbochargers and gas turbines, for nearly three decades, I'll try to give a bird's-eye view in a realistic manner and to offer a few worth-while comments of general interest.

Some Performance Features

Without repeating well-known thermodynamical matters, let us only briefly review what the addition of a heat exchanger (HE), Fig. 1, an intercooler (INT), and reheating (RH) can do for a simple, open-cycle gas turbine (SC) in regard to the amount of power produced per pound of air per second through-flow, and how the specific fuel consumption can be reduced. For gas temperature, heat-exchanging effectiveness, and component efficiencies, very reasonable and equal assumptions compatible with the present state of art underlie the curves shown.
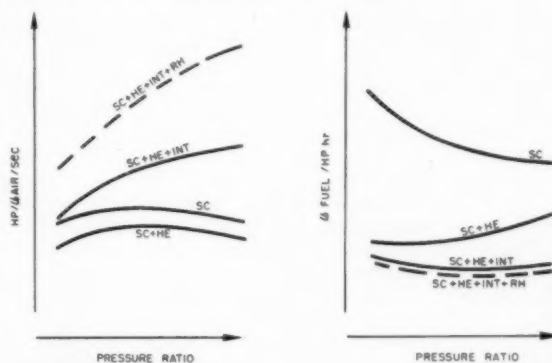
SC—simple, open cycle
SC + HE—simple cycle with heat exchanger
SC + HE + INT—simple cycle with heat exchanger and intercooler
SC + HE + INT + RH—simple cycle with heat exchanger and intercooler and reheat

Fig. 1 Specific power and specific fuel consumption versus pressure ratio for different gas-turbine cycles

The lowest fuel consumption shown corresponds to numerical values equaling today's medium diesel engines. The addition of air intercooling between compression stages lowers the compression work required while through reheating the expansion work can be increased; the cycle approaches more the Carnot cycle (1).[2] Although the opinion has been expressed (2) that this arrangement appears unlikely for vehicles, one may well debate this issue. For a given tip speed in the compressor, air intercooling leads to higher compression and expansion ratios which, with a given combustion-gas temperature, result in a lower temperature at the inlet to the exhaust-gas heat exchanger,

[2] Numbers in parentheses refer to the Bibliography at the end of the paper.
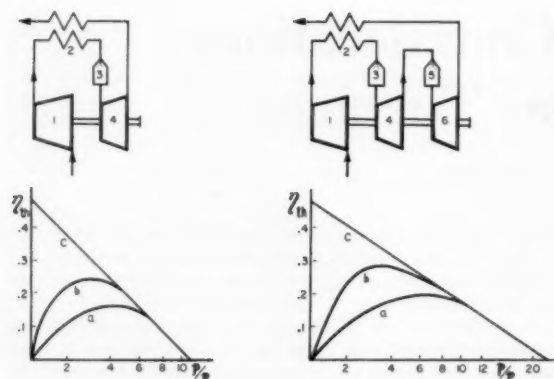
Fig. 2 Thermal efficiency of gas turbine versus cycle pressure ratio with (*a*) no heat exchanger; (*b*) with practical heat exchanger; (*c*) with infinitely large heat exchanger

thus not only simplifying its construction but also slightly reducing the significance of high heat-exchanger effectiveness. Naturally, intercooling appears simplest with gas turbines in naval installations where cool water is available. For certain applications though, air-to-air intercooling seems to deserve close consideration.

Eighteen years ago some of these advantages already had been pointed out (3) when gas-turbine development was in its infancy. Additionally, the influence of heat-exchanger size on thermal efficiency was demonstrated, Fig. 2. The calculations were then conducted for a 2000-kw unit with 1000 F gas temperature before turbine and standard day ambient conditions; intercooling between compression stages was not figured. Curves *a* represent conditions with no heat exchanger, curves *c* with an infinitely large heat exchanger, and curves *b* with some intermediate heat-exchanger size. The well-known trend towards higher pressure ratios required for decreasing heat-exchanger effectiveness is evident. In later Brown-Boveri gas-turbine development (4) which, incidentally, began in 1909 with the construction of a Holzwarth explosion turbine, intercooling between compression and reheating between expansion stages were used; better efficiency and substantially more power for the same air flow reduced the price per installed unit of power.

Before turning to matters of gas-turbine configuration and application, two more basic performance plots should be remembered. For a so-called split-shaft turbine, the dependencies of torque, fuel consumption, and power on output-shaft speed, Fig. 3 (5), plotted as abscissa, and with different gas-generator speeds (h—high speed; m—medium; l—low speed) are generally as shown. Numerical values naturally depend on the specific characteristics of the gas-turbine cycle and component efficiencies.

The characteristics of a gas turbine capable of delivering its output either in the form of compressed air (bleed flow) or in the form of shaft power, or in any combination thereof, and limited only by maximum rpm and gas temperature, Fig. 4, are as dependent on ambient temperature as other open-cycle performance characteristics. Power available, bleed pressure, and fuel consumption rise with decreasing compressor-inlet temperature, whereas, naturally, the bleed temperature rises with increasing inlet-air temperature.

For larger aircraft gas turbines, the two-spool arrangement has been introduced in jet and turboprop units (6). Dependent upon dimensions and controls, such a turboprop is capable of maintaining a certain power output from sea level to "full throttle" altitude, a characteristic comparable to turbocharged
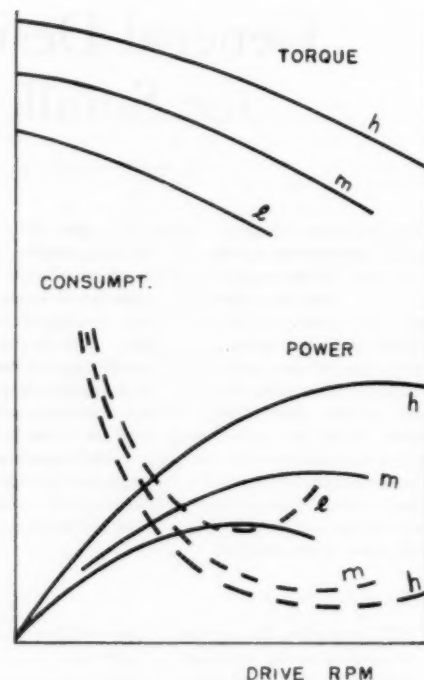
Fig. 3 Torque, power, and fuel consumption of split-shaft gas turbine for low (l), medium (m), and high (h) speed of gas generator versus speed of drive shaft
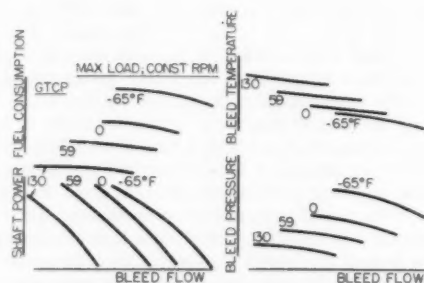
Fig. 4 Characteristics of gas turbine capable of delivering compressed air (bleed flow) and shaft power with air-inlet temperature to compressor as parameter

reciprocating engines. As the two-spool arrangement offers advantageous features, e.g., favoring maintenance of high compressor efficiency down to low pressure ratio, and since hollow-shaft arrangements already have been accepted for smaller gas turbines, it is quite conceivable to pursue the multi-spool configurations further also for the lower power class. Good fuel consumption possibilities are indicated and the design can allow for easy adaptability to more elaborate cycles.

Examples of Different Configurations

Before describing different configurations, let us summarize briefly the basic possibilities for smaller gas turbines.

Both with regard to compressor and turbine, the radial, the diagonal, and the axial type are being used, with the radial type generally favored. The diagonal-flow type, for which

quite some research and development time and money were spent, has only found limited application probably because the claimed advantages may have been found offset by certain manufacturing and/or weight problems.

Turbine and compressor casings are built with horizontal split plane or vertically flanged subdivisions and combinations of both. Rotor configuration, type of bearings, simplicity of manufacturing, and ease of servicing influence the decision where and how to subdivide casings.

While multistaging in axial compressors and turbines is quite common, multistaging in radial compressors is frequently considered somewhat less attractive and multistaging of radial turbines, especially with several radial-turbine wheels arranged in one housing, Fig. 5, appears not to have been considered sufficiently captivating.

For combustors, the can type with either straight-through or reversed flow, the multiple arrangement of cans in an annular space, the open-ring annular combustion chamber, and certain special elbow-shaped configurations are most common. Because of the advantages of cool-skin turbine housings, unitizing of so-called outer flame shell and outer turbine housing seems to become a preferred configuration.

The possibility of combining different basic component configurations allows for great freedom in arrangement and shape which may be particularly attractive for the lower-power-class gas turbines.

Since certain accessories must be integrated into a complete gas-turbine unit, the final choice of the configuration also can be influenced by the accessory-drive arrangement selected. Basically, all accessories can be mounted around the driving shaft either in a manner similar to the cylinders of a radial reciprocating engine or with all shafts parallel to the main engine shaft. The choice of the accessory arrangement necessitates different approaches to the problem of speed changes through gearing with spur gears of the straight or helical type and bevel-gear drives. So far as known, chain or belt drives are not in use with gas turbines in production. Although the high speeds of these turbine shafts offer difficulties, accessory drives without numerous precision gears appear attractive and possible. Of course, dependent on the application of the gas turbine, i.e., thrust producer, gas producer, shaft-power producer, or combinations thereof, and frequently also dependent on installation requirements, some of the designer's freedom in accessory arrangement vanishes.

Essentially, the component arrangement in small gas turbines can aim at the short type like a single-stage back-to-back arrangement or at the longer cylindrical shape as with multistage axial compressors and turbines, and annular combustion chambers. The smaller gas-turbine packages now produced in numbers for secondary aircraft power owe their shapes mostly to the necessity of squeezing the entire unit into the smallest possible installation space while keeping weight down, even at the expense of accepting restrictions in serviceability.

The following schematic illustrations are simplified. Some of them having great resemblance with existing small gas turbines, are, however, not meant to be a precise replica of actual designs; they are only indicative of interesting approaches.

In early work on gas turbines, both large and small, a single-stage radial compressor, can-type combustors, and axial turbines were somewhat favored. The housing and rotor arrangements of such a configuration, Fig. 6, resemble the exhaust-gas turbochargers used on U. S. fighter and bomber airplanes during the last war; this arrangement is rather simple both for a so-called turbojet or a shaft-power unit. More than one combustion chamber has been selected in better known gas turbines of this type such as those built by Boeing (7) and experimented with by
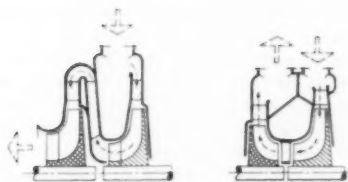
Fig. 5   Arrangements of two-stage radial turbines
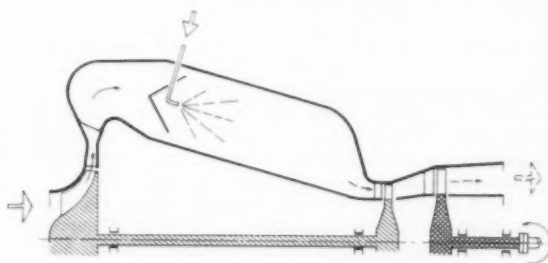
Fig. 6   Schematic of gas-turbine arrangement
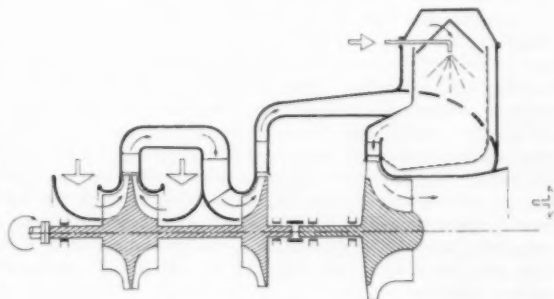
Fig. 7   Schematic of gas-turbine arrangement

General Motors (8). While offering easy access and good serviceability, as shown, this configuration is not best suited for smallest size, the length of the gas generator's rotor being largely dependent on the length of the combustion chamber. Integration of part of the combustor shell with compressor and/or turbine housing or using obliquely fitted-in arrangements are simple alternatives. And, when using more than one combustor, provisions for simultaneous firing (cross-over ducts) have been found necessary. The whole arrangement makes the individual development of each of the major components simple. The bearings can be inboard, i.e., with overhung wheels, or outboard, or with one wheel overhung and the other bearing outboard.

By combining features individually known to be advantageous such as high through-flow rates per reference area, a very broad compressor characteristic, i.e., flat dependencies between head and efficiency versus through-flow rate, as well as a centripetal turbine with cool-skin turbine housing, Fig. 7, a gas turbine can be enabled to serve as bleed unit, shaft-power unit, and the combination of both (9) without requiring complex protective controls. This approach furthermore allows for compactness while retaining individual subassemblies such as compressor, turbine, combustor, and accessories, simple to exchange in the field if required. While a single combustor has been shown with the discharge end of the combustion can fitting the inlet opening

of the turbine housing, multiple-can arrangements, a reversing flow, or an annular combustion chamber can be used. The cool-skin arrangement naturally reduces heat losses to the outside. If desired for reasons of simplicity, the blading of the three compressor-impeller sides may be identical. Whereas a fixed-shaft-type unit is shown, a floating stage arranged downstream from the turbine discharge can be added easily. If preferred, the rotor conceivably can be shortened with a single stiff shaft with two, rather than four, bearings.

In sketching for a simple gas turbine, the combination of a single-stage centrifugal compressor with a single-stage centripetal turbine doubtlessly appears very attractive mainly because of the simplicity of the rotor elements, Fig. 8. Naturally, a designer will try to shorten the length of the rotor as much as possible and, obviously, nothing prohibits letting the back sides of the radial wheels come in contact and finally become a single piece. With a combustion chamber of curved instead of straight shape connecting the compressor outlet directly with the turbine inlet, such unit represents the shortest simple arrangement.

Several small units with back-to-back rotors, Fig. 9, have been developed (10) with the major differences to be found in the choice of the combustion chamber, general casing, and the bearing arrangement.

Manifestly, straddle-type or overhung mounting easily can be used or one might even consider a semi-overhung type mounting with one bearing more or less hidden in the hub of the rotor, stress and temperature conditions permitting. The one-piece type of rotor is attracting a great deal of attention both in gas-turbine and exhaust-gas turbocharger work. Undoubtedly, there exists a not too simple problem of sealing the air flow against the gas flow at the outer circumference of the center disk. Yet what appears to be a weakness can be an advantage. A reasonable leakage rate from the compressor side over the rim into the turbine side can serve profitably as veil cooling over the hot face of the rotor. This feature is not being neglected in development work. The calculation of the critical speeds of this type of a rotor with long overhung mounting still represents quite a problem.

A unique approach to the combustion chamber and its fuel-injection system, Fig. 10, combined with a space-reducing double diffuser and two axial-turbine stages with a single-stage centrifugal compressor result in a rather compact unit (11). What normally would be the outer flame tube comprises the back wall of the compressor housing and the shell of the turbine casing, thus leading to a cool-skin unit. The rotor can be very rigid with outboard bearings on both ends, allowing operation below the critical speed if desired. Some portions of the combustion chamber are intriguing pieces of sheet-metal work in more ways than one. The basic components of this type have been developed in such a way that numerous types of gas turbines, turbojets, bleed units, single shaft, or floating power wheel could be offered by putting the building blocks (12) together in different ways, Fig. 11.

A simple rotor arrangement with two radial-type components, Fig. 12, combined with provisions for air filtering and sound attenuation at the compressor inlet and an exhaust-gas heat exchanger downstream of the turbine outlet with good exhaust diffusion, whose importance deserves full consideration, can be embodied in a compact arrangement (13) of cylindrical shape. Clearly, the rotor could be a single-piece, back-to-back arrangement. For instance, the rotor type shown would allow for a radial arrangement of the accessories required around the center portion of the main shaft.

A review of the more appealing arrangements would be incomplete without considering the radial compressor and turbine wheels with variable nozzle and/or diffuser area provisions,

Fig. 13 (14). Centrifugal compressors have been built with variable-area diffuser vaning in order to secure a wider and better delivery range. The centripetal-type turbine allowing for a very simple variable-area nozzle arrangement has been developed and is being used successfully, especially in aircraft secondary-power turbines. In addition to offering operation at higher efficiency with great variation in through-flow rate, such a nozzle arrangement in a centripetal turbine, as will be shown later, can be used advantageously for reversing and/or braking. Designing for the gas to leave the first-stage turbine wheel with high dynamic energy and properly shaping the turbine housing of the second stage, closely connected to it, for certain applications, may appear rather attractive, possibly in conjunction with a radial turbine containing no nozzle vanes. Although there seems to be no radial turbine with this characteristic on the market, the type is known to be under development (15). The illustration indicates two locations for combustion; i.e., reheat before inlet to the second turbine and also a heat exchanger downstream of the power turbine. Evidently, components similar to the ones shown can be arranged in many different ways worthy of further study.

Using the well-known combination of single-stage radial compressor and single-stage axial turbine in the generator portion, adding a floating power turbine and an exhaust-gas heat exchanger (recuperator), Fig. 14, has been judged an attractive arrangement (16). It would be fairly simple to design for a cool-skin turbine housing fully surrounded by the air discharged from the compressor. Instead of a single combustion chamber leading to a rather bulky protrusion around the circumference, subdivision into several small ones or an annular combustor appears possible. While at least one reversal in flow direction must be accepted, the heat exchanger can be of the rotary regenerator as well as of the recuperator type.

Instead of combining the single-stage radial and axial types, multistaging also can be used, Fig. 15. For a given pressure ratio, multistaging results in operation at lower tip speeds which, with a given gas temperature allows for usually cheaper construction material and can sometimes simplify obtaining a
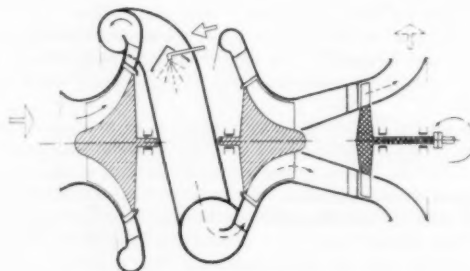
Fig. 8 Schematic of gas-turbine arrangement
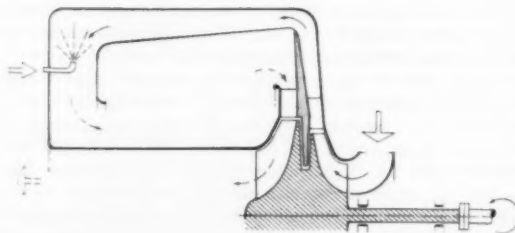
Fig. 9 Schematic of gas-turbine arrangement
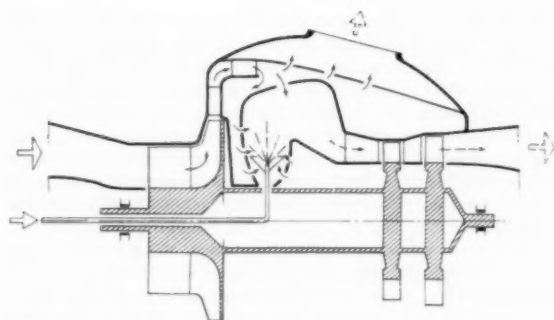
Fig. 10   Schematic of gas-turbine arrangement
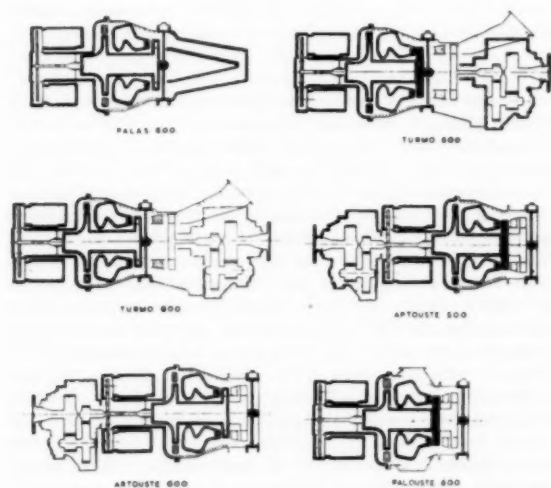
Fig. 11   "Building-block" possibilities with different components of Blackburn-Turboméca gas turbines (Courtesy Blackburn and General Aircraft, Ltd.)
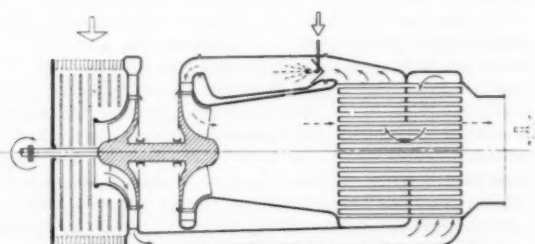
Fig. 12   Schematic of gas-turbine arrangement
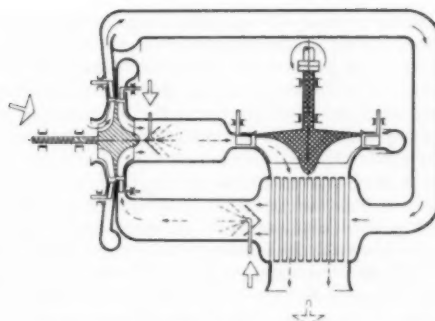
Fig. 13   Schematic of gas-turbine arrangement
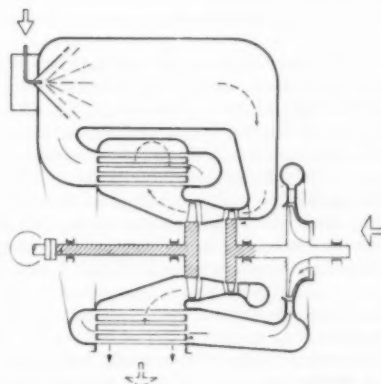
Fig. 14   Schematic of gas-turbine arrangement
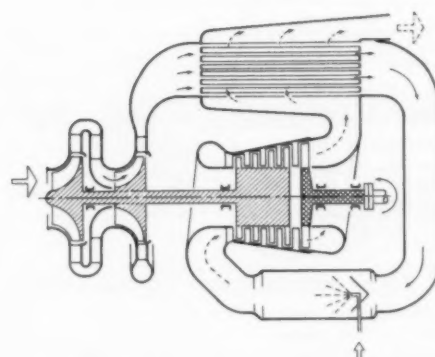
Fig. 15   Schematic of gas-turbine arrangement

good performance range. Also, lower rotational speed is advantageous with regard to the often required reduction gear at the power shaft of the turbine (17). In cases where quick acceleration of the unit is decisive, as in an automobile, the choice of the speed and moment of inertia of the rotor require a close study. Both with respect to combustion-chamber and heat-exchanger configurations, there exists a great freedom as long as both components are more or less separately attached to the outside of the unit casings rather than fully integrated. With basically the same components, other arrangements have been developed. If considered necessary, bearings downstream of the turbine discharge and surrounded by hot exhaust gas can be avoided in different ways, one of which is the utilization of a hollow-shaft arrangement (18). An arrangement with multistaged centrifugal compressor and axial turbines may be considered attractive, so long as instead of resorting to a heat exchanger satisfactory performance characteristics are to be obtained through using high-pressure ratio and gas-temperature values.

Where the gas stream leaving an axial compressor has to be bent around a corner, the centrifugal-type impeller practically suggests itself as the last compression stage, Fig. 16. If length of the complete gas turbine is more significant than diameter, a reverse-flow arrangement in the combustion chamber can be selected (19). Various bearing arrangements including the hollow shaft, as shown, can be adapted to this basic turbine configuration; their choice will greatly depend on the power take-off requirements as well as installation considerations, thrust utilization of the exhaust gas in aircraft installations, and so on.

Where fuel, such as natural gas, is plentiful and cheap, a heat exchanger often cannot be justified economically. If preference is given a setup with a single combustor, Fig. 17, or, in other words, when, as in some stationary application,
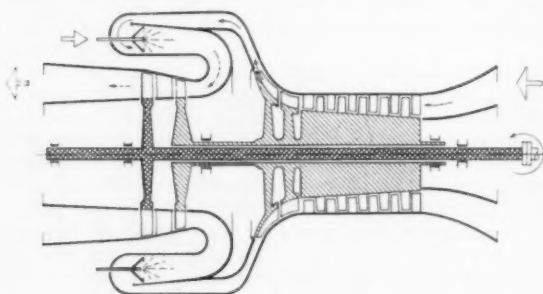
Fig. 16 Schematic of gas-turbine arrangement
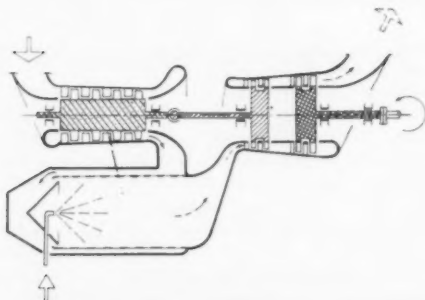
Fig. 17 Schematic of gas-turbine arrangement

space is not at a high premium, a simple arrangement can be selected and so arranged as to offer incorporation of heat-exchanging equipment (20) when operational conditions so demand. With a straight axial configuration of the rotors the total unit length can be kept short at the expense of reversing the flow before entering the combustion chamber and of increasing the diameter of the powerpack by arranging the combustor around the compressor casing. Again, depending upon the desired location of the power take-off, hollow shafting (21) can be chosen.

Finally, a special arrangement, rather than configuration, called "differential gas turbine," should be mentioned (22). Output shaft, compressor shaft, and turbine shaft are interconnected in a differential gearing. The advantages claimed for this arrangement are greater turbine power at low output speeds, when compared with the split-shaft arrangement, fuel savings, improved compressor acceleration, and the possibility of torque reversing to make the power unit act as a brake.

Applications

The fact that for the past few years gas turbines with up to around 300-hp output are being produced in volume clearly demonstrates their acceptance. The financing of the development of small gas turbines has found support from Air Force and Navy because this power plant is particularly well suited as a source of secondary aircraft power for electrical or hydraulic services, starting, pressurization, climatization, and the like, delivering useful output in the form of shaft power and of compressed warm air and/or hot gas. The compressed air is fed into pneumatic starters and other air-turbine-type drives including air-cycle climatization. Shaft power is needed for alternator and pump-drive means so that electrical and hydraulic energy are available without operating the main power plant. The gas-turbine type offering delivery of compressed air, if necessary with temperature selection, and, at the same time, shaft-power output, has found good acceptance because of its adaptability to the demands. The success of these turbines (23) is largely the result of the selection of components with broad operating capabilities and the perfection of a so far unsurpassed degree of automaticity of the control system, giving true push-button operation. Small gas turbines (24) have been installed both in the aircraft and under it in so-called pods for take-off assistance or secondary power and on many airfields they are available in ground-support equipment; i.e., wherever the advantages of ease of starting, of light weight, small size, adaptability, independence from cooling water, and so on, really count, while fuel consumption, in fact, is of lesser importance. Thousands of these units are in daily operation and have justified themselves beyond question. As a matter of fact, even for the highly specialized purpose of just serving as a starter for large gas turbines, they are being used, for instance, on the 12,000-shp turboprop engines powering the Russian long-range bomber *Bear* (25).

The large gas turbines for aircraft propulsion, of which approximately 100,000 units have been built in the United States of America alone after World War II, have progressed so well that now they have been accepted as the most promising power plant for new airliners. This, as well as the success of the small secondary power gas turbine, explains why smaller gas turbines are finding more and more application as prime movers for winged aircraft and helicopters and are being installed as take-off aid (26), source of power for boundary layer control, and so on.

While so far the aircraft-type gas turbines in production range from about 50 hp upwards, smaller units may well become of interest in conjunction with special projects. The desirability of a gas turbine below, say, 20 hp may be considered questiona-

ble. Unless unorthodox design approaches are developed, the creation of such a small gas turbine to become a success is no picnic, although definitely not impossible especially if fuel consumption is rather insignificant.

The pros and cons of small gas turbines for road vehicles have been debated at length in the press. Meanwhile, a number of financially strong manufacturers of reciprocating automotive engines have published early results (27) of their activities in the development of gas turbines in the 100 to 500-hp class.

It is probably safe to agree with statements by others that the gas turbine will first be introduced in trucks and similar heavy-duty equipment, although presently, no one will say when. "Such vehicles can be designed to exploit the full advantages of the turbine in its various configurations. These advantages include (a) reduction of weight and bulk, (b) use of available low-cost fuels, (c) operation in any climate, and (d) low maintenance" (28).

Some day the event of producing gas turbines in "Detroit quantities" can be expected. In trying to assess the chances of the small gas turbine, the enthusiast cannot afford to ignore further progress in the development of reciprocating automotive engines. If the same degree of optimism regarding further perfection is afforded both types, the future reciprocating engine for the automobile may give the competing small gas turbine a hard run for the money. Too many intangibles make any clear-cut comparison presently impossible. Nevertheless, where true experience is available, such as the admirable record of the Rolls-Royce Dart gas turbine in commercial airline service, the following statement (29) cannot be overlooked:

"The day-to-day inspection and maintenance of the gas turbine in military and in airline service absorbs far less man-hours than does piston engine inspection. Present day experience still confirms this.

"It is, therefore, fortunate that the gas turbine lends itself better to more rapid flight development and acceptance than the piston engine."

Naturally, to draw a fine line between small, medium, and large-size gas turbines is impossible, and there are applications where a size differentiation has little meaning. There are others where this kind of a prime mover, so flexible in design and arrangement, can utilize widely different components enabling it to compete for almost every small prime-mover application. These features are attractive, and sometimes deceptive. The situation reminds one somewhat of reciprocating engines where who knows how many engines with wobble plates, sleeve mechanisms, and what have you have been invented all right but have found little practical application.

Certain interesting possibilities may be worthy of a brief discussion.

All-electrical transmissions are being used on diesel-electric locomotives and ships because they offer torque amplification without gears, braking capability, easy reversing, finely adjustable speed, agreeable acceleration characteristics, and eliminate drive shafts and related equipment. A more recent example for such a system was the Le Tourneau "Sno-Train" which is powered by d-c motors driving from inside the rims through reduction gearing. Converters serve to use d-c motors on a-c power generation. Plainly, it suggests itself to consider a gas turbine as the prime mover of the a-c alternator or, in other words, a smaller and more compact electrical power package. For alternator drives, the so-called single or fixed-shaft gas turbine is quite suitable and has been shown to give satisfactory service in hundreds of lightweight a-c power sets, a typical example being the GTP70 developed and manufactured by The Garrett Corporation and in service, in quantity, in Navy-assigned aircraft.

Testing of small gas turbines with mechanical drive in a locomotive (30) has produced promising results such as: "The gas turbine mechanical-drive combination will outperform a diesel-electric of the same engine-output rating."

For instance, the versatility offered through the utilization of existing gas producers of the turbine type was recently expressed (31) as follows:

"There is a great deal to be said for using high production-rate jet engines as replaceable short-life gas-producing sections and developing longer-life power turbines and gears for the output section. This approach is satisfactory if the basic aircraft engine is reliable within its operating life at the derated naval rating. Engine development cost would also be a great deal less by using aircraft gas producers."

Small gas turbines are quite suitable as drivers for turbotype pumps such as widely used for irrigation purposes. Where smudging is required around citrus plantations the turbine exhaust may represent a good source of heat. Portable fire-pump sets, gas-turbine driven, have been introduced successfully. The hot exhaust gas from gas turbines has been found to offer excellent ice and snow-removal possibilities.

Turbochargers of certain types now being produced in great numbers have amazingly low price tags. Now add what is necessary to make such a turbocharger operate as a simple gas turbine without shaft-power demand. When produced in sufficient numbers and with the simplest solution for combustion chamber, fuel and oil pumping, and so on, the price scarcely can be expected to remain an obstacle against using such an utterly simple unit as a hot-gas producer. Operating with an air/fuel ratio above 70:1, the contents of burned gas in the air are so low that such a heater might well be acceptable for certain purposes. Especially if such a unit is hand-crank started and requires no battery and no electrical gadgetry, it could easily take over where heaters now used are useless because they depend on electrical energy. One can stretch the imagination still further. Such a unit conceivably can satisfy three purposes and offer an economical system when needed.

1 It can preheat a reciprocating engine, or the engine compartment, before starting and then, directly or indirectly, serve as starting means.

2 After the reciprocating engine is in operation, the compressor and turbine of this gas turbine can function as the engine's turbocharger bypassing the now unnecessary gas-turbine components.

3 Supply power boost through additional burning in the gas turbine's combustion chamber within the temperature limitation of the turbine.

Since a gas turbine is capable of certain services more easily than other prime movers, it can be expected not only to compete with those but to open a new market where such special features may command even a premium in initial cost. Power-plant size and weight cause concern in offshore oil drilling and pumping. Artificial islands are not particularly cheap and, if due to complete absence of vibration the foundation requirements could be lowered a great deal, such feature should prove attractive. Needless to mention that the well-known greater freedom in the choice of fuel can offer worth-while advantages. In gas-line pumping, large gas turbines have earned a good reputation. Certain thoughts about smaller-size gas turbines for such gas-line pumping have been offered. For collecting the gas from numerous smaller fields into a large system, the convenience of a simple, small, portable prime mover with no cooling water requirements is interesting. Its natural characteristic of increasing power output with decreasing ambient temperature seems to meet pumping requirements nicely while using the gas

piped as fuel. In sewage-disposal establishments, the organic gaseous fuels produced can be burned in gas turbines which, in turn, serve as compressor drives or for electrical power, or similar services.

In conjunction with blast-furnace plants, the combination of bleed and shaft power in one gas turbine seems to offer attractive features. When projecting the gas turbine as a power plant in process cycles, the particulars of such process cycles must be studied closely in order to utilize fully the flexibility of gas turbines, mainly where different kinds of liquid or gaseous fuels are plentiful and where the normally lost exhaust heat can be recovered (32). For use in so-called undeveloped areas, portable power sets for pumping, current generation, heating, or cooling, and capable of using a wide variety of fuels, offer advantages. A small gas-turbine-powered generator set could be carried over distances on the back of a mule or a camel and set up ready for operation in a very short time, whereas the now-existing equipment is more cumbersome.

While mentioning examples for applications, the closed cycle as developed by Escher Wyss and the semi-closed cycle as used by Sulzer Brothers should not be overlooked. The latter one appears to have the benefits of the closed cycle, however, without requiring all the cooler equipment.

Although, apparently not in use, a gas-turbine cycle which might be called "subatmospheric" appears worth mentioning. If hot waste gas at atmospheric pressure is a given source, it may be burned in a turbine driving a compressor which lowers the turbine exhaust pressure below atmospheric and recompresses the gas to be discharged into the atmosphere, and, in so doing results in net drive power available. Naturally, there must be a heat exchanger lowering the gas temperature between turbine exhaust and compressor inlet, both for thermodynamical and mechanical reasons.

Development Problems

General. Before turning to some development matters of the major gas-turbine components, the old question regarding the desirability of smaller aircraft propulsion gas turbines (33) may justify a few observations. The so-called cube-square law based on the fact that, when maintaining essential similarities, thrust obtainable is an area function, i.e., diameter squared, while weight changes with volume, i.e., diameter cubed, indicates advantages for the smaller unit well worth closer examination. Although true in principle, there obviously are practical limits resulting from manufacturing and handling concerns. And there are others. To maintain a certain pressure ratio and only reduce the rate of through-flow involves a change in the so-called specific speed of the turbomachinery components which eventually will indicate a change in rotor-component configuration, i.e., eliminate geometrical similarity as a basis of comparison. Difficulties in scaling down result from the combustion chamber where geometrical similarity might reduce the chamber length below an acceptable minimum. So, within "reasonable" limits only can this law be applied as has been demonstrated to a degree by units such as the Rolls-Royce Soar lightweight turbojet engine. The very low specific weight of approximately 0.15 lb per pound of thrust, as compared to about twice the specific weight of big engines, and the high figure of over 1300 lb thrust per sq ft frontal area show that a cube-square law "relates change in engine dimension with engine weight" as a particular engine design is scaled up or down. There appears to be a practical size limit for the small turbojet engine expected by some to be around, say, 2000-lb thrust, by others around 50 to 70 lb per sec air flow, beyond which practical necessities invalidate the attractiveness of the small turbojet. However, figures of $^1/_{10}$-lb engine weight per pound of thrust are

not considered just a utopian dream. Some numerical values published forecast a thrust-to-weight ratio in excess of 15:1 if currently undertaken advanced research work, such as supersonic compressor operation, good turbine blade cooling, and so on, is completed successfully. A recent study suggests that characteristics particularly attractive in the 2 to 3 Mach-number range can be obtained in smaller simple gas turbines without sacrificing performance and offering a worth-while weight advantage. One turbine stage and 4 to 5 stages in the compressor with a total pressure ratio between 4 and 7 in conjunction with presently acceptable gas temperatures offer good performance up to a Mach number of about 3. For certain propulsion applications, notably VTOL aircraft, "power plants of extremely low weight . . . . . . with acceptable cruise fuel consumption" (34) are more or less a matter of success or failure.

When speaking of gas turbines, people are inclined to think of the screaming and "fuel thirsty" jet engines powering military aircraft with wide-open compressor inlet and turbine outlet, taking air in and discharging hot gas at very high velocities. These observations, however, must not be interpreted to mean that a gas turbine is noisy *per se*. Why should a gas turbine be much noisier than its older brother, the steam turbine? Ducting and installation influence the noise level decisively. One of the most disagreeable contributors to noise is the high-pitched scream emanating from an open compressor inlet. Whereas, not too long ago, this fact was considered rather detrimental to the acceptance of smaller gas turbines, it can be stated safely that this difficulty has been reduced sufficiently. As indicated in some of the schematic figures, rather narrow parallel channels formed by splitters, linings with sound-absorbing fabrics, and properly adapted air filters at the compressor inlet really can do a silencing job. A recent demonstration proved that a high-speed gas-turbine power set easily could be made to satisfy the aeromedical noise specifications; whereas the presently used reciprocating engine does not.

Compressor. Among the major components of a gas turbine, a good compressor is considered the most costly and time-consuming matter of development. The unqualified statement that the efficiency of an axial compressor is superior to a radial compressor has been repeated too often. Compressor designers know that far too many factors to be listed here enter a comparison. Aside from the specific speed and configuration relationship, qualified engineers probably agree that the design criteria for axial compressors, into whose development much more effort has gone, are somewhat more thoroughly established than those for small high pressure-ratio radial compressors. By nature, the radial compressor offers broader performance characteristics, lesser manufacturing headaches, and, last but not least, a robustness and insensitivity to fouling that an axial compressor will scarcely ever have. Just the same, both types can be expected to share "the market" since each has certain inherent advantages and disadvantages.

Remarkable progress has been made in small radial compressors (35) and experimental results reported lead us to look forward to single-stage centrifugal compressors with tip speeds of 2000 fps and more, giving upwards of 6:1 pressure ratios under standard day, sea-level conditions with efficiencies above 0.75. Rolls-Royce single-stage centrifugal compressors—at least experimentally—small and large have been operated with pressure ratios above 5 for many years. For gas turbines, with emphasis on power concentration, high component efficiencies are decisive. If, at a pressure ratio of 4, the compressor efficiency can be raised from 0.75 to 0.85 and the turbine efficiency from 0.8 to 0.9, the power output will increase by almost 75 per cent. Supersonic flow must be dealt with in high pressure-ratio single-stage compressors. The development of suitable diffusers becomes

more and more critical and may lead to somewhat unorthodox approaches. Having demonstrated impeller-tip speeds above 2000 fps to be safe almost 20 years ago, one can hardly be accused of undue optimism in expecting the progress in materials to allow for higher tip speeds now or in the future. Just as it is good practice to avoid large center holes in highly stressed disks, preforging with proper grain size and directional grain-orientation control are significant. The double-sided symmetrical semi-shrouded impeller configuration should be of advantage in regard to stress problems.

Whenever intercooling between compressor stages can be used, high-pressure ratios are most easily obtainable at moderate stresses, both mechanical and thermal. Assuming two centrifugal-compressor stages with, say, 2.75 pressure ratio each and intercooling to within 30 F of ambient temperature, a value allowing even air-to-air intercooling, the total pressure ratio is about 7, i.e., of a magnitude usually high enough for economical smaller gas turbines of high specific power production. At present, small gas turbines for auxiliary power operating around 3:1 are in production. While the provision of an intercooler of high effectiveness can be objectionable for aircraft turbines, the author believes it to be worthy of further studies even for mobile equipment and especially so in naval vessels. Admittedly, the added complexity and increased plant size must be weighed carefully; yet, too numerous to be listed here appear the various advantages. And, again, with design ingenuity, one can expect to find solutions wherein through fullest exploitation of the flexible adaptability of the turbo-unit to multiple functional requirements (36) the specter of complexity can be banished.

Turbine. There is no need to discuss "axial" versus "radial" turbines. The radial type has found wide acceptance for gaseous substances only within the past decade, with the axial type its senior by about 40 years. Small radial turbines are being used by the tens of thousands and have established a remarkable record for themselves. Regarding price, manufacturing simplicity, ruggedness, and so on, the thoughts mentioned in conjunction with radial compressors are equally or even more strongly applicable to the centripetal turbine.

The outstanding simplicity of nozzle arrangements with variable throat areas and the possibility of reversing appear unchallenged. Several hundred radial-type turbines for air or gas with variable-area nozzles are in use. So, the practicability of this attractive and economical type of power control is no longer a matter of doubt. A properly designed variable-position nozzle ring permits reversing the direction of rotation of the turbine wheel. Work on this feature has been or is being conducted in several places. Recently published results (37, 38), reported to have been obtained in actual tests conducted on an 11-in-diam wheel, reveal that, with a maximum efficiency in the "forward" direction of rotation of approximately 0.86, reversing the direction of rotation through changing the angles of the nozzle vanes produced an efficiency of approximately 0.57 at about the same velocity ratio as in forward operation and with constant pressure ratio, Fig. 18. A study of such a reversing-type turbine compared to a reversing-gear arrangement or a reversing propeller for ship propulsion is said to have shown the reversing turbine to have by far the lowest relative cost, the lowest relative weight, and a relative efficiency only slightly lower than with the two other arrangements (38). In some cases the nozzle vanes can be eliminated altogether, a welcome simplification since the development of nozzle diaphragms with the inherent tendency towards cracking is no minor task.

Because of the high density of the working substance in a closed cycle, the dimensions of a turbine can be kept small; as an example, one radial-turbine wheel of about 28 in. diam operating at 10^4 rpm is quoted to be good for 10,000 shp. Whereas,

in this case, the diameter ratio selected is about 1.9, typical for fairly low specific speed, development work in recent years has proved that good performance can be obtained with wheel OD to outlet diameter ratios below 1.5, which means successful application up to specific speeds in a range less desirable with axial wheels.

Almost everyone who has worked with hot turbine wheels operating at high stress and temperature, and subject to frequent and rapid starting and shut down, remembers thermal cracks around the rim of the wheels as one of the difficult problems, especially when the combustors tend to torch. Although perhaps not quite as drastic, rim cracking, in some cases, also has caused extra development work with radial-turbine wheels of the semi-shrouded type. Stationary tests simulating thermal shocks by alternately heating and quenching full disks, after only a few cycles, have produced small cracks similar to those observed after a number of accelerated starting cycles and rapid load changes with early gas turbines. After extending the starting cycle from 30 to about 90 sec, surface cracks did not occur. Surveys under operational high-temperature conditions have indicated gradients of approximately 50 F per in. in the central portion of a small radial wheel and about 10 F per in. in the outer portion of the rear shroud or disk tested. Calculations suggested a change in thickness and taper of the disk to reduce the total stresses. A slotted wheel rim has been proposed. Whether any reduction of the surface stresses upon sudden temperature change would be effected is uncertain. Changes in the shape of the rear shroud and/or the elimination of a part of it, somewhat comparable to the oldest radial impellers of the "star type," proved successful, offering, at the same time, a welcome means of lowering the axial thrust. Fully shrouded wheels may be chosen in cast, riveted, or brazed construction, moderate speed and temperature conditions of operation permitting.

It is known that an exact stress calculation for the most common open-face type impeller is a difficult and very time-consuming task even when the radial temperature gradient is small or negligible; and it is almost equally difficult to measure the actual temperature gradients, varying greatly dependent upon load and/or speed. Fortunately, in radial-turbine wheels with predominantly accelerated flow and less danger of flow separation, liberties in blade shaping can be taken. Where the stresses are most critical, aerodynamical compromises seem not to be unduly detrimental; and, where the aerodynamical shaping is somewhat critical, the stress problem can be met more easily.
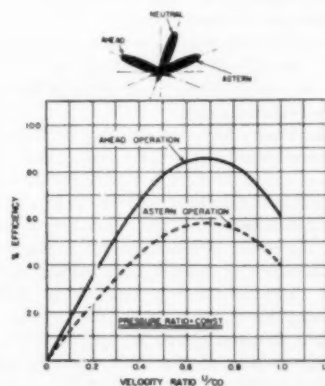

Fig. 18 **Efficiency of reversible centripetal turbine as function of velocity ratio. (Results reported by Escher-Wyss.)**

All these observations are part of the reason for the author's almost proverbial optimism regarding further extension of the utilization of radial-type turbines. Excellent performance up to over 1600 F and 1600 fps tip speed has been demonstrated now in thousands of units over adequate operational times and in spite of the fact that the time elapsed between pushing the starter button and full power availability is below 30 sec.

Since about 1939, every now and then research results regarding blade and disk cooling for axial turbines have been published (39). The advantages of operation at lower stress levels, i.e., increased safety, or at higher gas temperature must be weighed carefully against increased design complexity and it is thus hardly surprising that little agreement on this subject can be found among engine designers both with respect to air and liquid cooling. By increasing the gas temperature of a modern turbojet unit from 1500 to 2500 F, a thrust increase of around 50 per cent is conceivable without permitting higher metal temperature in critical turbine elements. The better known methods of blade cooling are convection cooling, veil or film cooling, spray cooling, and transpiration cooling. The degree of complication considered acceptable and design ingenuity paired with advanced manufacturing techniques are main factors in the evaluation of the different cooling approaches. For the Rolls-Royce Dart turbine, extensions of the turbine-blade roots have been developed (40) which give a considerable reduction of temperatures at critical points, such as lowering the rim temperature from approximately 1065 to 700 F.

Except for results (41) obtained by blowing cooling air over the back face of a radial-turbine wheel, no data including blade cooling seems to be available publicly. The statement that heavy cooling may prove to be more difficult in a simple centripetal-turbine wheel is probably descriptive of the present state of the art; nonetheless, partly hollow rotors and veil cooling appear feasible.

In mentioning development in the direction towards higher performance, an interesting presentation (42) indicating possibilities with lower rather than higher working temperatures for aircraft gas turbines has been made. In general, resort should not be taken to higher working temperatures until all other methods of improvement have been exhausted, especially when such methods mean greater safety, longer life, greater freedom in materials, and numerous other significant advantages.

Combustor. While most gas-turbine components for the normal, constant-pressure cycle can be designed on the basis of rather scientific analysis without too much "hit or miss" activity, the design and the development of combustion chambers still are more in the category of art than science.

Major requirements are as follows:

(*a*) To burn efficiently one given (or several different) fuel(s) without smoke and without depositing carbon.

(*b*) To be easy to ignite and not to flame out upon sudden flow and air/fuel ratio changes.

(*c*) To operate with high heat release without localized hot spots and with well-equalized outlet temperature.

Combustion chamber loading is commonly expressed in heat units per volume unit, time unit, and referred to a pressure unit. High turbulence is essential for short flame length, good temperature profile at the outlet of the combustion chamber, and high heat release. Differently shaped flame holders, swirl caps, and similar means produce different degrees of turbulence and naturally influence the pressure loss, desired to be small. The specific loading is proportional to the approach velocities of the air and inversely proportional to some characteristic dimension (43). For equal air-mass flow and air/fuel ratio one single chamber can be shown to require more space than several geometrically similar small chambers and these may even

operate at a lower pressure loss. Because of the dimensioning influence of how fuels are prepared (44), i.e., gasified, vaporized, or atomized (45), it is impossible to make general statements for liquid, gaseous, or pulverized fuels. For gaseous fuel, interesting results from combustion-chamber research have been reported (43). A spiral chamber, a torus-shaped chamber, and a so-called return-flow chamber have been studied, suggesting possibilities of unorthodox arrangements in turbines in that some of the repeated deceleration-acceleration flow processes conceivably can be eliminated. Permitting an engineer's imagination to wander a little, flow "mysteries" where free vortex flow changes into solid body flow together with the flow conditions at rotating compressor and turbine cascades may lead to very simple arrangements reviving the thinking about "rotating compressor-combustor-nozzle assembly" proposed (46) in conjunction with a floating power unit many years ago in Germany. In the small combustors tested, astonishingly high heat releases were measured and flow conditions maintained in which the flame never came in contact with the walls of the combustion chamber, floating like a cloud in the center portion of the chamber. Heat-release rates of more than ten times the values now realized in operating gas-turbine combustors have been found to be obtainable through simple means and with much improved pressure-loss conditions.

Whereas in earlier days and for stationary units refractory liners or liquid-cooled walls have been used, such complications appear unlikely with smaller gas turbines which benefit from the extensive aircraft-turbine development. One finds the straight-through type combustor, the so-called elbow type and multiple-injection annular types, just as the fuel can be admitted in numerous different ways from high-pressure injection through simple nozzles to spill-type nozzle to prevaporizers, and other preparing means. Naturally, fuel injection by itself, particularly with low-grade liquid fuels of high viscosity, can pose serious problems. The chamber has to fulfill the requirements of separating primary and secondary air properly, of discharging uniformly, and of forcing the air to flow in such a way as to avoid local overheating, both on the inner liner and the outer casing. The statement that the three factors—temperature, turbulence, and time—must be well co-ordinated is probably as brief and good a rule as can be given. And, in observing the major different schematic arrangements for small gas turbines, the combustors are conspicuous by their space requirement in comparison to other components.

Owing to the higher fuel consumption of simple gas turbines the burning of cheap fuels is always as attractive economically as it can be troublesome for the designer. The utilization of natural, sewage, and blast-furnace gas offers few major difficulties, whereas with residual fuels the three main problems—combustion, corrosion, deposits—are as yet not resolved satisfactorily. To use "undoped" fuels appears fairly hopeless; fuel additives such as soluble silica material, powdered Kaolin, magnesium, zinc sulfate, and others were mentioned in the 1955 London Conference on Combustion. The true working mechanism regarding deposits and counteraction with dopes is probably still unknown. In conjunction with the development of more powerful aircraft-propulsion units, the designation "exotic" fuels has been used frequently in recent months. Little can be published on this subject at present. "HiCal is a combination of boron, carbon, and hydrogen...... Boron is the most efficient carrier of hydrogen and is lightweight" (47). Pentaborane (48), a liquid hydride of boron, has a heat content of 29,127 Btu per lb and weighs 38 pcf, offering 16 per cent more Btu's per cu ft than kerosene. The practical introduction of such fuels should not be expected soon.

Bearings, Accessories, and Miscellaneous Items. Antifriction

bearings or journal bearings? That is one of the tough questions the gas-turbine designer faces too. Both types are in operation and are likely to remain, especially when both are given the benefit of better mounting or supporting arrangements as used, for example, in European exhaust-gas turbochargers (49) and the Boeing gas turbine (50), i.e., resilient centering and improved lubricant feeding.

Sometimes bearing problems seem to have been overrated in magnitude. The reasons for their representing a development problem, especially with smaller gas turbines, are mainly higher sliding velocities combined with space restriction and heat influx. Many small high-speed turbomachines even with elevated temperatures of the working substance are functioning well; most of them use the highest precision antifriction bearings and sometimes tricky lubrication schemes. Race-curvature studies and bearing preloading, work on new types of separators and separator materials, and the like, are consuming plenty of efforts although an assessment of the real difficulties often reveals that they were created with the design. The operational conditions, for instance, in small gas turbines just as in turbochargers, make it mandatory to balance dynamically the rotating groups within very fine accuracy lest the bearing loads exceed acceptable limits. Then, in field operation the rotor goes "out of balance" for any one of innumerable reasons. One inch-ounce unbalance in a rotor operating at 40,000 rpm will cause a bearing loading approaching 3000 lb. The close clearances of antifriction bearings, which certainly offer many nice features, in conjunction with the unavoidable although small manufacturing tolerances combined with excessive rotor loads will excite vibrations which, in turn, can destroy the bearings quickly and promote fatigue cracking in other structural components. Similar considerations also are applicable to journal bearings. In both cases, the problem can be reduced significantly if, instead of forcing the rotor to rotate on the drawing's center line, it is permitted to rotate around its real c.g. The practical solution to this is known as resilient bearing mounting which within built-in limits allows the rotor to seek its own c.g. because the bearing is not supported rigidly. For antifriction bearings various solutions have been worked out such as semi-tangential leaf springs supporting the bearing shell or a spiral-type spring surrounding the bearing shell or support by a fluid cushion. With journal bearings resilient mounting has been used too and the so-called floating sleeve with its additional oil film has been found to be successful not only for journals but also for thrust bearings (51). Aside from mechanical load, heat influx, especially upon sudden shutdown from full load, often causes trouble in bearings located adjacent to hot wheels or in hot shells. Heat dams resulting from either thinned-down metallic connections or from using coatings or materials with very low conductivity are usually resorted to.

Although oil, fuel, air, and gas seals belong to the somewhat cumbersome items, space unfortunately excludes detailed descriptions. Suffice it to say that face-type seals have been developed for higher rubbing velocities and pressure differentials and that for certain types the sealing pressure can be co-ordinated with the turbine operation by tapping into the best suited sources of balancing medium such as air. For stationary gaskets, soft metals, and newer materials such as plastic or teflon-type vellum sheet and combined structures containing steel, asbestos, and so forth, are offering a wide choice.

As far as accessories are concerned, gas turbines generally require pumping means for fuel and lubricants, starting and ignition equipment and, dependent on their special duties, an electrical generator, governing and control equipment, and filtering provisions. The accessories represent a sizable item with smaller gas turbines; on one of the smallest aircraft turbo-

jet engines the weight of such accessories exceeds 130 lb representing about one third of the total engine weight. Without going into further details, attention is invited to a few features favoring the so much needed simplicity especially on smaller gas turbines. Fuel pumps of a delivery characteristic such as to avoid excess fuel flow into the combustion chamber and, thus, eliminating the need for nozzle-box or tail-pipe temperature sensing are feasible. Most gas turbines today still depend on electrical energy for starting and ignition, a feature that is sometimes detrimental to the acceptance of a small gas turbine. For units of apparently up to several hundred horsepower output, manual starting has been demonstrated. Also, flame-lit combustion chambers have been shown to be possible. While for large gas turbines pneumatic starting, hot or cold, is now commonly used in aircraft applications and electromotor or auxiliary turbine starting for stationary gas turbines, the process of diminution can hardly be expected to go so far as to build tiny little turbomachinery-type starters for smaller gas turbines. Pneumatic starting, however, can be considered if semipositive displacement-type starters were developed. Of course, cartridge starting with direct impingement on to the rotor is a known means.

Combined with the development problems of smaller and smaller accessories is that of the gear train connecting the accessory drives with the high-speed gas turbine shaft directly or after a speed-reducing power-gear train. It is not known whether belt or chain drives are being used with success.

The demands on automaticity for smaller secondary aircraft-power gas turbines have been very exacting and have led to the maximum thinkable—namely, single push-button operation regardless of output demand and ambient conditions. In certain aircraft where small gas turbines drive electrical alternators the frequency and paralleling requirements are just as critical as in large power stations and speed control within $1/_{10}$ per cent has been specified together with speed recovery below one sec after big load changes. Naturally, a high degree of automaticity with all the added equipment can hardly be expected to increase the functional reliability and in many cases product improvement work had to be carried on at great expense because of accessory trouble while the gas-turbine unit proper operated well.

Although space limitation does not permit going into details, a general remark regarding unit design may be in order. The build-up or arrangement should be such that adjacent components maintain their relative positions independent of temperature variations; temperature gradients in heavy sections should be kept to a minimum; changes in thickness should be gradual and never abrupt. Casings which have to stay on center can be kept in line by radially free three-point mountings.

In closing it is deemed timely to mention that, in spite of the multitude of components, a gas turbine can have significantly fewer parts and a greatly reduced number of critical tolerances than now known reciprocating engines capable of operation under comparable conditions.

Heat Exchangers. Although presently there appears to be no exhaust-gas heat exchanger in quantity production for smaller gas turbines, great efforts are being made, especially in conjunction with the development of automotive-type gas turbines (52), to find a satisfactory solution. Generally, the names "recuperator" and "regenerator" are used to describe heat exchangers. The recuperator is characterized by the fact that walls always separate the cold and the hot heat-exchanging gases, whereas in a regenerator cold stream and hot stream go through heat-absorbing "matrices" alternately. Whereas a recuperator has no moving parts, there are always some, even if only simple ones, in a regenerator installation.

The two more important configurations for stationary heat
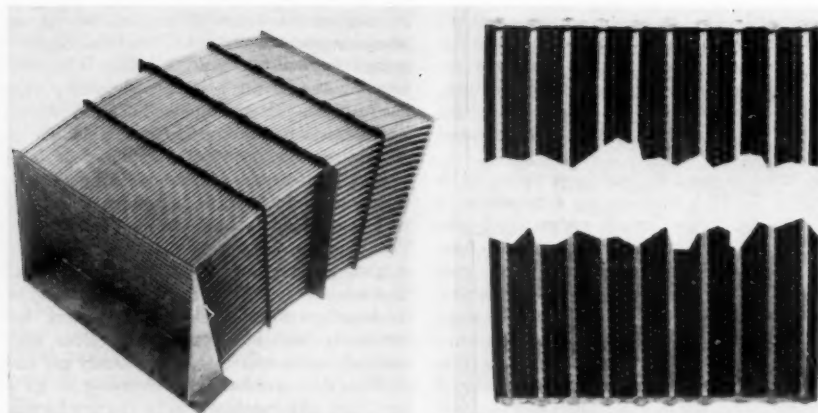
Fig. 19 Tubular and plate-fin type heat exchangers. (Manufactured by The Garrett Corporation, Los Angeles, Calif.)

exchangers (recuperators) are the tubular type and the plate-fin type, Fig. 19. Heat exchangers of this kind are presently used in many aircraft applications for lowering the bleed-air temperature of the main engine. Such stainless-steel plate-and-fin heat exchangers of a high effectiveness for airborne applications have been developed with stainless-steel plates, nickel fins, and copper brazing for maximum temperatures up to around 850 F.

Basically, to transfer heat from, say, gas to air is a simple problem so long as economy is not too much of an issue. Heat exchangers for gas turbines, unfortunately, demand best economy both with respect to effectiveness and pressure loss and space requirements, and, to make matters worse, of course, regarding materials and production. Through analysis of heat-transfer and flow-friction relationship, one finds that for a given problem statement small hydraulic diameters lead to small heat exchanger space requirement, relatively speaking, which is one of the reasons of the increasing importance of the plate-and-fin type in general. In comparing the two types shown, one might be justified in saying that, whereas development problems may be somewhat simpler for the tubular configuration, in mass production the plate-fin type may be preferred. Studies about repairability and practical field experience may alter this observation. The plate-fin construction appears to lend itself to the building-block idea. Because of packing more surface into a given space, its weight appears not to differ greatly from the tubular type. In gas-turbine applications especially for mobile equipment the problem of dirt accumulation and fouling are of great significance. The more surface is packed into a given volume, the more likely dirt accumulation can become a grave problem. While not a great deal has been published about recuperators for smaller gas turbines (53) and all attention appears absorbed by the regenerator, it cannot be overlooked that a cross-flow type recuperator often will remain a good contender.

The original idea of a heat-storage material, the matrix, arranged in a rotor for heat-exchanging purposes, dates back many years to Ljungström's air preheater (54) design and was actively pursued for gas-turbine application by Ritz in Germany about 20 years ago. All along, disk-type and drum-type matrix arrangements have been considered. With proper guidance of exhaust gas and combustion air to and from, the continuous rotation results in alternate heating and cooling of the matrix. Obviously, the same effect can be obtained in other ways. The fact that no such regenerator has as yet reached a true production perfection indicates that "all that glitters is not gold." Indeed,

there are some genuinely challenging difficulties of which the sealing problem is probably the most effort-consuming (55). To prevent loss of compressed air as well as possible and at the same time to avoid too much seal friction seem to be almost incompatible requirements. Optimal dimensions can be established for matrix flow length, frontal area, rotational velocity, and the optimum ratio of cold to hot frontal area; the seal leakage matters are still somewhat obscure. Although remarkable progress has been made and astonishingly high effectiveness values between 80 and 90 per cent have been claimed, numerous problems especially in regard to manufacturing methods remain to be solved.

Among the various proposed regenerative systems, the "turbo-regenerator" (56) may deserve further study. It "combines its main function of heat exchange with that of power output......" Essentially, it is proposed to use the heat-exchanging gases to drive the rotor and thus eliminate the problems of the mechanical drive means such as gears. The stator of such unit contains a system of nozzle vanes while the rotor comprises moving blading with the flow deflection so arranged as to exert the torque required to overcome the mechanical and flow-friction losses.

The author was fortunate in obtaining information about an interesting type rotary heat exchanger from one of the early pioneers (57). This "twin piston regenerator," Fig. 20, consists of two identical pistons, a center housing, and two identical end casings which contain one matrix each. The operation, as described by the originator, is as follows:

"This regenerator consists essentially of five castings; i.e., two identical end casings, a center casing, and two identical pistons.

"The end casings each contain one readily removable matrix, consisting of alternate layers of corrugated and plain metal strip loosely wound round a core. This core also provides cylinders, in which the pistons slide.

"The center casing is subdivided by a vertical wall, into two halves each containing one cylindrical sealing face, an annular gas duct and, in a common flange, a gas inlet or an air outlet respectively.

"The pistons have the form of tubes with enlarged centers. They carry three pairs of piston rings and have in the center a spider forming a boss to which the operating mechanism is connected.

"The operating mechanism can consist either of a mechanical, periodically operated lever or, as shown in the drawing, of two interconnected hydraulic cylinders.

Fig. 20 Twin-piston regenerator. (Courtesy C. A. Parsons & Company, Limited, Dr. H. L. Ritz.)

"The hot gases enter the center casing through the rectangular opening in the top flange. With the pistons in the position as shown, the gases are directed towards the near end cover, pass through the matrix where the main part of their heat content is transferred and stored and leave cooled through the LP-piston and the far end cover.

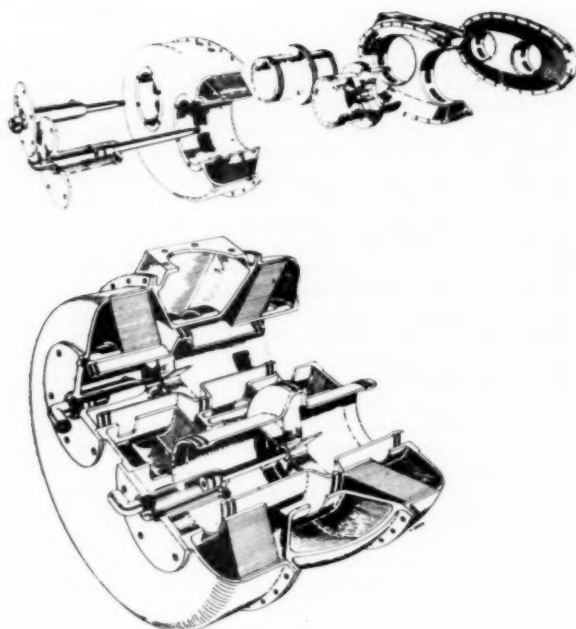"The cold compressed air enters the heat exchanger via the second opening in the far end cover and leaves, after having picked up heat in the matrix at that end, the center casing through the round opening in the top flange. When the heat stored in that matrix is exhausted, the pistons are reversed simultaneously so that the hot gases reheat the cooled matrix and the cold air flow via the HP-piston through the matrix heated in the previous period. In this way, the hot gases and the cold air pass alternately and in counterflow through the two matrices releasing and absorbing the heat to be exchanged.

"While flowing along the pistons the cooled gases and the cold compressed air pass through the spiders, the radial webs of which act as cooling vanes and together with the increased turbulence in the enlarged center sections keep the hot sealing faces on a temperature level which allows for the application of standard materials and lubricants. The clearance on these center faces is such that only the sealing rings slide, the pistons being entirely supported by their cold end sealing faces."

Design and functional characteristics are described in this way:

"This heat exchanger, by virtue of the temperature symmetry in axial direction and the existence of one temperature level only in each cross section, is free from distortion. The moderate temperature and small loading of the sliding faces practically exclude wear. The axial gas pressures on the pistons balancing each other, the operating force is negligible. Due to the ideal sealing contours the leakage is very small. Corrosion or partial burnout of the matrix does not result in breakdown as the surfaces transferring the heat do not simultaneously seal the high

pressure air. For the same reason the matrix can consist of inexpensive, thin material.

"Small hydraulic diameters with high heat transfer and consequent economy in the use of matrix material can be realized since no brazing of the corrugated strip is required. The auto-cleaning effect of the periodic gas reversal and expansion eliminates the danger of choking of the small passages while the loose matrix can be readily unwound for cleaning. The low matrix weight has the essential advantage of a small thermal inertia.

"This version of the piston regenerator has been developed for incorporation into a conventional, open-cycle gas turbine with a can-type combustion chamber arranged alongside the turbine casing. The top flange of the heat exchanger is attached directly to the exhaust outlet of the turbine and to the air inlet of the combustion chamber, the only gas duct needed being that for the cold compressed air from the compressor to the heat-exchanger end cover."

For certain gas-turbine cycles so far not being actively pursued for the smaller gas turbines, namely, the closed-cycle (38) and coal-fired gas turbines, heat exchangers are no longer a matter of choice but rather a prerequisite, and one may expect further contributions from these activities as well as from the development of nuclear power plants (58).

Future development and testing are expected to shed light on the relationship of control problems and acceleratory ability of a gas turbine, and the volume of "stored" gas and air in heat exchangers.

Materials and Manufacturing. The future of the smaller gas turbine depends at least as much on the development of better materials as that of the big units and possibly even more so in that the cost of materials and simpler ways of shaping them may be a more critical issue when smaller units become a mass-produced item for industrial and automotive power. Basically, for the components of the turbine proper, structural materials of high strength at high temperature with satisfactory resistance against scaling and corrosion and—for the rotating parts—the best possible ductility are required. Good elongation is essential because in highly stressed rotating parts, particularly with asymmetrical shapes, it allows for some stress equalization. There are alloys of good strength and ductility at elevated temperatures, yet at room temperature they are too brittle and difficult to handle. Where different alloys compose a subassembly the matter of differential thermal expansion can influence the design or the choice of material. Wrought nickel-base alloys, e.g., Inconel X and Nimonic, are widely used for blades and wheels including semi-precision forgings when the operational conditions, i.e., high stress and temperature, so dictate. Of course, much more "modest" alloys, such as 19-9DL, 19-9DX, 18-8, and so forth, are well proved and suited for normally stressed hot wheels. Among the more recent nickel-base alloys, Inconel 700 is considered strong and useful up to 1650 F; it is an age-hardenable nickel-cobalt-chromium alloy with good resistance against corrosion and oxidation; it can be forged, machined, and welded. With 45 per cent nickel, 28 per cent cobalt, and 15 per cent chromium, its availability might not always be unlimited. Inconel 713C is a nickel-chromium cast alloy with strength up to 1700 F, good resistance to thermal fatigue and, because easy to cast, it is considered attractive for gas-turbine blades. It contains no cobalt but 1 to 3 per cent Cb + Ta. As investment-cast material, it is used in the as-cast condition without subsequent heat-treatment; it is not available as wrought material. The strength and elongation properties (59) at higher temperatures, especially at 1700 and 1800 F, are attractive. In sheet form, materials such as Inconel 702 and Incoloy 901 are fairly well known. Inconel 702 appears to have a particularly good oxidation resistance and in tests con-

ducted between temperatures of 2000 and 2400 F Inconel 702 was better in oxidation resistance than Inconel which, in turn, is superior to Inconel X. Incoloy 901 has properties comparable to much higher alloyed materials yet it is low in strategic alloy since it represents a modification of the German wartime Tinidur. It has been recommended, for instance, for turbine disks and other components requiring good rupture and creep strength in the range of 1000 to 1400 F.

As is well known, the properties of high-temperature alloys are affected by the melting procedure, the fabrication, and the heat-treatment used. And, since higher operating temperatures will always be attractive for high-performance gas turbines, the metallurgist is continuously searching for better superalloys. As has been reported (60), vacuum-induction melting and vacuum-arc-melting practices are promising further progress. The higher rupture strength of vacuum-melted alloys is partly attrib-uted to the possibility of higher titanium and aluminum contents. Rupture stress, rupture time, and elongation have been improved. Especially the improved ductility obtained appears important since, often, ductility was assumed to decrease when raising the rupture strength by conventional means such as heat-treatment or increased alloy content. The general conclusion reached is that vacuum-melted alloys display marked superiority with respect to elevated-temperature strength and ductility over corresponding air-melted material. Numerous alloys have been tested to allow for a comparison of the mechanical properties in the range from 1000 to 1600 F (J-1570; Waspaloy; L-605; S-816; 19-9DX; GMR-235; N-155; and so on).

Alloy R-235 (61) is produced by vacuum melting; it is a nickel-base, aluminum and titanium-containing alloy with good properties up to 1750 F. Attractive features are its low strategic-alloy content, good oxidation resistance, and resistance to over-aging in service. Its fabrication properties are considered good and the cold-working characteristics have been called exceptional for such strong alloy. Newer cast alloys are appearing on the horizon. The possibilities of using precision-cast high-tempera-ture alloys can be assumed to be familiar. Cast individual blades and complete blade rings are being used in axial turbines and compressors. Precision-cast gas-turbine exducers as well as complete one-piece radial wheels, Fig. 21, have been accepted widely especially in exhaust turbochargers with inlet-gas tem-perature up to around 1400 F. Whereas only a few years ago a

Fig. 21 Typical shapes of mass-produced radial turbine wheels—forged and cast

material thickness, say, at the trailing edge of a blade around 0.02 in. appeared difficult to achieve, precision castings with such dimensions are now produced successfully in large quantities. In many cases, the surface finish of the as-cast piece is quite adequate without further refinement. With precision castings the design freedom is increased; blade or cavity shapes too difficult or too costly for machining and/or forging easily can be employed. Naturally, the progress in precision casting repre-sents a challenge to develop new forging and extrusion techniques. Recent work has made it practical and economical to precision forge complete radial-turbine wheels in high-temperature alloys with the radial blades requiring no machining. Although this technique is still fairly young, optimism about further progress appears justified. A general comparison between precision forging and precision casting appears rather impossible at this time because the techniques are still being improved, the designs are still somewhat in a state of flux, and the number of wheels produced per time unit certainly influences the cost picture (62).

As was somewhat implied, the author still has some misgivings about cast hot wheels. Whereas with the forged wheel the soundness of the forging can be inspected to a high degree through, for instance, ultrasonic methods (63) where unacceptable discontinuities within the material reflect the sonic beams as "echo signals," an equally reliable way of checking the soundness of a casting has not come to his attention. X-ray inspection alone has its known drawbacks; gamma-ray inspection appears more satisfactory. Hot spin-testing of each hot wheel of mass-produced turbines scarcely can be expected to become common practice and would still leave doubts. Naturally, the degree of concern greatly depends on circumstances which are difficult to assess. Since turbine wheels usually give little or no warning of impending failure, protective provisions of some sort may be deemed indispensable. Exploding automotive torque converters are bad enough; an exploding turbine wheel can be more catas-trophic. Where weight and space and cost conditions permit, a "solid" casing capable of unconditional containment can be an acceptable yet not always simple solution. This matter of containment is attracting a great deal of attention partly origi-nated by some recent rather unpleasant events in airplanes. The concern for human lives and/or very expensive equipment make a definite solution to the containment problem a must. For aircraft installation the "armor" approach with its salient weight increase is not attractive. For axial turbines the use of a "frangible" turbine wheel has been advocated and shown to offer a possibility (64). Although possibly slightly more complex, turbine wheels equipped with aerodynamical provisions limiting the maximum possible rpm safely present a promising solution with a small weight penalty. Developments in this activity are too young to suggest a summarizing statement. While the safety measures listed may be of great importance, entire protec-tion is still not given against a perhaps unlikely yet not impossible premature wheel burst at normal operational conditions. At any rate, quality control may not often have been as much a matter of protecting human lives as it is in hot, high-speed turbomachinery. Extensive creep-rupture investigations of materials must be continued.

Before leaving hot-wheel matters a word about a related item: One design and manufacturing problem doesn't appear to be too drastically different with different turbine-wheel configurations, namely, the attachment of the wheel to the shaft. The oldest and simplest method, of course, is the disk with the center hole fitted to the shaft with or without shrink and with or without provisions to allow for flexibility in the radial direction. Addi-tionally, one-piece forgings, i.e., wheel and shaft in one piece, have found wide application. The development of such forgings with prescribed grain flow has given the forging people quite
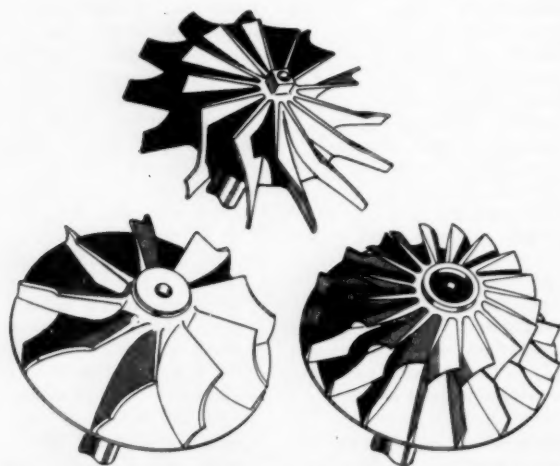
some problems, but it appears that these are essentially matters of the past. Numerous other methods of connecting shaft and disk solidly through different types of welding or brazing, as well as with bolt connections of the flange-on type of the center-bolt type including "Curvex" or "Hirth" self-centering members, are being used. Whenever stresses are high, special attention must be paid to the elastic deformation of the disk in rotation in order to eliminate out-of-balance conditions resulting from a shift of the center of gravity. Engineers who have developed this type of turbomachinery often have experienced rather severe vibrational problems and traced the excitation back to out-of-balance rotors. This problem is not new and some leaves could be taken from old steam-turbine publications.

High-temperature brazing (65) is becoming more and more significant in the construction of gas turbines. Real progress has been made in the recent past indicating that brazing techniques will allow for greater freedom in using complex structures, in joining cast to wrought material, in allowing for combining rather different metal thicknesses, and also in joining dissimilar metals. Attention is invited to the problem of sometime detrimental interaction where certain dissimilar materials are joined and may operate in unusual surroundings.

As mentioned, operation at much higher gas temperature may become feasible through component cooling. For stationary items such as nozzle vanes nonmetallic materials are being considered and developed (66). Ceramic materials can be made but their safety has been called low. Metal-ceramic blades are more promising in terms of high-temperature strength and creep and shock resistance. "Kentanium" (67) has been used in an experimental hot turbine of the radial type for 100 hr at 1850 to 1950 F at interesting tip speeds. This material is produced from pressed and sintered powdered components on the titanium-carbide basis. Cermets, in general, are finding increasing attention. Metal-bonded carbides (68), borides, oxides, and nitrides are being studied. Low ductility, high notch sensitivity, and low impact resistance have to be improved greatly.

Among the most interesting materials for compressor components and certain casings, titanium seems to rank highest, its strength characteristics approaching some steels now used, e.g., in disks, yet with only about 60 per cent of the density of steel. Lower blade weights produce lower rim stresses, then allow for reduced disk dimensions. Thus the weight advantage easily can be higher than the density ratio would indicate. Decreased rotor weight can permit reduced housing requirements and is very beneficial for reducing the moment of inertia or, in other words, improving accelerative ability of a rotor. Presently, titanium is still a fairly costly material. Its increasing utilization in aircraft gas turbines, big and small, and the development of more and more facilities for titanium production are likely to remove this obstacle eventually to make the utilization of its light weight, good resistance to corrosion, desirable strength-to-weight ratio, and fatigue-strength to tensile-strength ratio feasible in commercial power units.

Meanwhile, aluminum in cast or wrought condition is being improved continually and remains a favored material especially for compressors. Sintered-aluminum products (68) made from fine aluminum powder with 10 to 20 per cent of aluminum oxide by a process of combining cold and hot pressing and extruding have been reported to show promise for use at temperatures of 800 F and possibilities of varying strength and ductility properties "by changing the amount, size, and distribution of the oxide phase." Needless to say that the designer will always wish for improved magnesium alloys and read with delight about the application of certain cast magnesium alloys "to have resulted in burst speeds in excess of 2000 ft per sec" (62) in compressor development work.

Closing Remarks

With all the large gas turbines for aircraft propulsion in stationary plants and also as locomotive drive, where one company alone has some 40 units of the same type on order and/or built, and/or in operation, quite some experience in operation, maintenance, and reliability has been accumulated and—so to speak—has helped to establish good ground rules. With smaller gas turbines the same is only true so far as military use is concerned. For all the special purpose units as well as in automotive use experience is nearly nil, thus, any extrapolation is risky. Following are random examples of progress:

A good demonstration (38) was recently given in a comparison of a 2000-kw closed-cycle power unit. The 1939 test unit weighed 25 tons and was 32 ft long; today's version is listed at $7^{1}/_{2}$ tons and is 10 ft in length with the number of wheels and blades of the new plant reduced to a small fraction in less than two decades. Rolls-Royce Dart gas turbines, propelling Vickers-Viscount airliners, have demonstrated more quality features than almost any piston engine of equal "age" and under comparable conditions. Seemingly, this unit matured before it reached its 'teens! A Metropolitan Vickers 1750-kw gas turbine-generator set at Venezuelan La Concepcion oil fields, burning natural gas, has been in service since mid-1955. Boeing engineers have stated that the 502 gas turbine "has conclusively demonstrated that life is superior to many reciprocating engines" and that "overhaul life and cost in production quantities are approaching reciprocating engines . . . ." "Our design approach to competition with reciprocating engines, then, is to combine simplicity and durability to produce a long-life engine which can be built, operated, and maintained economically in a variety of applications without significantly impairing the other inherent advantages of gas turbines—compactness, light weight, favorable torque characteristics, and vibrationless running " (69).

Accepting all favorable comments at face value and placing all the confidence in the ingenuity of our engineers and manufacturers, we still must not allow undue enthusiasm to run away with us. Today's reciprocating engines are by far not the last word for this type of power plant. Recently, one of the country's leading engine men predicted a reciprocating car engine approaching 1 lb per hp and power ratings of 400 or 500 hp from engines not exceeding today's package size. In comparison, the GMC 304 automotive gas turbine inclusive of heat exchanger and gearing has been reported to have a specific weight of about 4 lb per hp. The ratio of specific weights of smaller gas turbines of comparable power but basically different design type appears to be as high as 1:35 (GE T58, Clark Mark TA). That rugged simple construction and engine life satisfactory for a good many stationary applications (5000 hr) can be obtained at considerably lesser weight penalty is exemplified by a recent newcomer, the British Standard Motor Company 250-hp gas turbine (10s), the specific weight of which appears to be below 2 lb per hp. This figure is within the range of the specific weights of the majority of auxiliary mobile smaller gas turbines and indicates that design simplicity can go far in diminishing the otherwise appreciable weight difference between limited-life and long-life machines.

Especially with diesel engines, through the acceptance of turbocharging and the development towards high boost pressure and charge intercooling, the thermal efficiency is climbing nicely. A good deal of sound attenuation results automatically from turbocharging. Torque characteristics are being improved steadily. Air-cooled car and truck engines, both diesel and Otto types, are well established. Promising developments for diesel engines capable of using lower grade fuels are under way. Thought is being given to better integration of auxiliary service equipment such as heating, cooling, and braking. New ways of moving the valves indicate possibilities for faster running

engines. And there are many more real improvements in the making.

Now, let us look at the smaller gas turbine. Remarkably low fuel consumption has been shown; yet, it isn't a really simple gas turbine any more. As a matter of fact, it may not be easy to meet the size and weight figures just mentioned. Although not as critical as a diesel engine, so far the gas turbine cannot be called ready to thrive on really cheap or low-grade fuels. Engineers qualified to judge have stated that they expect to find retaining of automatic transmissions, comparable to, although somewhat simpler than, the ones now in use, desirable or necessary in conjunction with gas turbines, thus, disillusioning simplification expectations. And, last but not least, presently a gas turbine is still likely to require more of the so-called critical materials than the reciprocating engine. All of this probably means that the introduction of the gas turbine as a common vehicle drive should not be expected in the immediate future. The tremendous investment in tooling now in existence, and mostly greatly specialized, cannot be ignored. Quite a bit of new tooling will be required for the economical mass production of gas turbines. Additionally, to provide for country-wide service people and facilities is no minor task either. Some day it will happen because there are many good reasons for optimism. Intensive efforts towards the perfection of smaller gas turbines must continue. As pointed out earlier, this power producer is capable of doing many things better now than the reciprocating engine. And as usual, the greatest need appears to exist in conjunction with equipment for supreme military forces.

Once fully developed, a gas-turbine power package should be cheaper to manufacture and better in service. Its flexibility and almost vibration-free operation are valuable assets just as its ease of starting in lowest ambient temperatures and the modest mounting requirements. Further improvements in power concentration, i.e., horsepower per pound of air per second, are foreseeable and can be expected to become mandatory. And, finally, equipment rather comparable to gas turbines as we know them now should be of great usefulness in conjunction with power production from nuclear energy.

In closing, the author wishes to express his sincere appreciation for material made available by many co-operative friendly engineers in United States and foreign industries.

Bibliography

1 "Gasturbinen," by K. Bammert, *VDI Zeitschrift*, Bd 97, 1955, pp. 1039–1043.

2 "Vergleichende Betrachtungen über Antriebsquellen für Schwerlast-Kraftwagen," by F. Nallinger, *VDI Zeitschrift*, Bd 97, 1955, pp. 233–241.

3 "The Combustion Gas Turbine: Its History, Development, and Prospects," by Adolf Meyer, Proceedings of The Institution of Mechanical Engineers, vol. 141, 1939, p. 205.

4 "A Further Stage in the Development of the Combustion Turbine," *The Brown Boveri Review*, October, 1956, p. 263.

5 "Boeing's Model 502 Gas Turbine," by H. M. Jacklin, Jr., *SAE Journal*, vol. 58, August, 1950, pp. 53–57, 72.

6 "Bristol Orion Narrows Gap Between Turboprops and Jets," by S. G. Hooker, *Aviation Age*, April, 1957.

7 The well-known Boeing 502 series gas turbines have been tested in a wide variety of applications and have been in production for some time: "The 200-HP Boeing Model 502 Gas Turbine," by S. D. Hage, *Interavia*, vol. 4, June, 1949; also footnote, 5.

8 "Firebird II: New Gas Turbine Powered 'Dream Car' Designed for Highway Use," by W. A. Turunen and J. B. Bidwell, *General Motors Engineering Journal*, vol. 3, March–April, 1956, pp. 43–45.

"G.M.'s Regenerative Engine—Compact Automotive Power Unit With a Twin Drum-Type Rotary Heat Exchanger," *The Oil Engine and Gas Turbine*, vol. 24, December, 1956, pp. 318–320.

9 This approach has been selected by The Garrett Corporation; thousands of such units have been delivered and are being used for secondary aircraft power installed in the aircraft fuselage or in pods

or in ground carts. See *The Oil Engine and Gas Turbine*, vol. 24, October, 1956, pp. 237–239.

10 The Solar Aircraft Company's small gas turbines, known as the Mars series, are being built in quantity and have established a good record.

(a) See "Radial Flow Compressors and Turbines for the Simple Small Gas Turbine," by L. F. Wosika, Trans. ASME, vol. 74, 1952, pp. 1337–1347.

(b) "A New Solar Gas-Turbine-Driven Auxiliary Power Plant," by R. Kress, ASME Paper No. 56—GTP-8.

(c) "A Pod-Mounted Gas-Turbine-Driven Auxiliary Power Unit," by R. Kress, ASME Paper No. 56—S-10.

(d) "The Small Gas Turbine: Problem and Promise," by Sumner Alpert, SAE Paper No. 30, 1957.

(e) More recently back-to-back rotors have been shown by W. H. Allen Sons & Co., Ltd., *The Engineer*, vol. 197, May 14, 1954, pp. 709–711; Budworth, Ltd., *The Oil Engine and Gas Turbine*, vol. 24, December, 1956, pp. 316–317; and Standard Motor Company, Ltd., *The Oil Engine and Gas Turbine*, vol. 24, March, 1957, pp. 431–435. In all three examples simplicity and ruggedness have been emphasized.

11 This concept underlies all Turboméca gas turbines known today. Gas turbines of this type are being produced and widely used as turbojets, fixed or split-shaft power units, and as bleed units. *Flight*, vol. 59, February 1, 1951, pp. 135–136 and "La turbine à gaz, moteur de demain," in *La Technique moderne*, Tome XLII, Nos. 5/6, 1950, pp. 80–88. Licensees: Blackburn and General Aircraft, Ltd., *Flight*, vol. 67, June 3, 1955, pp. 753–757, and Continental Aviation and Engineering Corporation.

12 The Blackburn redesigned units known as 500 and 600 series embody certain modifications aimed at improvement in the production aspect. The 500–600 demonstrate the flexibility by building units up from the basic unit which is the jet-type engine.

13 "Parsons Gas Turbine for Moderate Powers," *The Oil Engine and Gas Turbine*, vol. 22, February, 1955, p. 387.

14 Patent Specification No. 1,018,006, Centrax Power Units Limited and Richard Henry Howard Barr.

15 Miehle-Dexter Turbochargers; *Diesel Engine Catalog*, vol. 21, 1956–1957, p. 280.

16 "Rover's Auxiliary Power Unit," *The Aeroplane*, vol. 87, November 12, 1954, pp. 698–701.

"Britain's Latest Turbine Car," *The Oil Engine and Gas Turbine*, vol. 24, October, 1956, pp. 231–232.

17 An arrangement basically like the one shown has been selected for the Austin Turbine Motorcar engine.

"The Austin Vehicle Gas Turbine," by J. H. Weaving, ASME Paper No. 57—GTP-2; and *The Oil Engine and Gas Turbine*, vol. 24, February, 1957, pp. 391–395.

18 A configuration of the type shown was selected by the Italian Fiat Company, *Gas and Oil Power*, January, 1956; "Development of Gas Turbines for Road Vehicles," by A. Carelli, ASME Paper No. 57—GTP-3.

19 "Some Experiences in the Development and Application of Lycoming's T53 Gas Turbine Engine," by Anselm Franz, SAE Paper No. 32, 1957; and *Aviation Week*, March 25, 1957, p. 40 for T55.

20 Ruston Hornsby TA type industrial gas turbines—"New Gas Turbine," *Automotive Industries*, vol. 115, September 15, 1956, pp. 60–61, rated at 1000 kw without and 750 kw with heat exchanger. A somewhat comparable arrangement although with entirely different combustor design is the 1000-kw marine auxiliary gas turbine by W. H. Allen Sons & Co., Ltd.

21 Bristol Proteus turboprop engine, *Flight*, vol. 56, August 18, 1949, pp. 191–196.

22 "Differential Gas Turbine," by D. W. Hutchinson, *SAE Journal*, vol. 64, June, 1956, pp. 70–75.

23 GTC85 Series, a product of The Garrett Corporation; see reference (9).

24 The best known examples are gas turbines by: See 9, 10, 11, and Boeing Airplane Company "Evolution of a Gas Turbine," by W. E. Skidmore, SAE Paper No. 31, 1957; Continental Aviation Engineering Corporation: "Continental's Small YJ69 Turbojet," *Aviation Age*, May, 1955, p. 34; and "The J69 Engine—Two Applications of a Small Turbojet," by D. J. Todd, ASME Paper No. 56—SA-62.

25 *Aviation Week*, April 8, 1957, p. 53.

26 "The Future of Small Turbojet Engines," by A. T. Gregory, Fairchild Engine Division paper presented at Symposium on "The Role of the Small Gas Turbine," October 1, 1955, New York, N. Y., sponsored by Polytechnic Institute of Brooklyn.

27 In the United States:

(a) The Chrysler Corporation: "The Automotive Gas

Turbine—Today and Tomorrow," by G. J. Huebner, Jr., paper presented at the SAE Meeting, October 8, 1956;

(b) General Motors Corporation: "The Regenerative Whirlfire Engine for Firebird II," by W. A. Turunen and J. S. Collman, SAE Paper No. 772, 1956;

(c) Ford Motor Company: "An Exploration of the Automotive Gas Turbine," by A. H. Beaufrere, SAE Paper No. 437, 1955; also, *SAE Journal*, vol. 63, March, 1955, pp. 66–69.

In Europe:

(d) The Rover Company, Ltd.: "Rover Gas Turbine Car," *Gas and Oil Power*, vol. 51, October, 1956, pp. 233–234;

(e) The Austin Motor Company, Ltd.: "The Austin Turbine Motorcar Engine," *The Oil Engine and Gas Turbine*, February, 1957, p. 391; and "The Austin Vehicle Gas Turbine," by J. H. Weaving, ASME Paper No. 57—GTP-2.

(f) Fiat: "Development of Gas Turbines for Road Vehicles," by A. Carelli, ASME Paper No. 57—GTP-3.

28 "The Role of the Turbine in Future Vehicle Powerplants," by C. G. A. Rosen; the second L. Ray Buckendale Lecture, 1956.

29 "Birth of an Engine," by Air Commodore F. R. Banks, *The Aeroplane*, vol. 92, March 15, 1957, pp. 366–367.

30 "Development of the First Gas-Turbine Mechanical-Drive Locomotive," by E. L. Barlow, Jr., *Mechanical Engineering*, vol. 77, 1955, pp. 133–136.

31 "Design Considerations for Naval Gas Turbines," by G. L. Graves, Jr., ASME Paper No. 56—GTP-1.

32 "The Small Gas Turbine Cycle and Exhaust-Heat Recovery," by A. A. Hafer, *SAE Journal*, vol. 64, October, 1956, pp. 37–38.

33 (a) "Why Small Engines?" by D. P. Edkins and M. H. Thorson, SAE Paper No. 29, 1957.

(b) "Choice of Design for an Advanced Turbojet," by F. H. Keast, *Canadian Aeronautical Journal*, vol. 2, November, 1956, pp. 322–328; also, *The Aeroplane*, vol. 91, August 10, 1956, pp. 189–191.

(c) "The Pattern of Tiny Engines," *The Aeroplane*, vol. 92, January 11, 1957, pp. 52–54.

(d) "Potential Advancements in the Field of Small Aircraft Engines," by R. V. Hensley, S. R. Shapiro, and Raymond Capiaux, SAE Paper No. 82, 1957.

34 "Powerplants for VTOL Aircraft," by P. G. Kappus, Institute of the Aeronautical Sciences Paper No. 667, 1957.

35 (a) "Design and Development of a Broad-Range, High-Efficiency, Centrifugal Compressor for a Small Gas-Turbine-Compressor Unit," by I. E. Speer, Trans. ASME, vol. 75, 1953, pp. 395–407.

(b) "Development of a Centrifugal Compressor for a Small Gas-Turbine Engine," by N. R. Balling and V. W. Van Ornum, ASME Paper No. 56—GTP-2.

(c) "The Austin Vehicle Gas Turbine," by J. H. Weaving, ASME Paper No. 57—GTP-2.

36 An example coming to mind is the combination of an axial turbine with an axial blower in one wheel whose blower blades represent the spokes. This very old idea was recently found to offer preferred characteristics in an air-conditioning system.

37 "20,000-SHP Nuclear Propulsion Plant," by R. P. Giblon and G. H. Kurz, *The British Motor Ship*, vol. 37, March, 1957, pp. 518–522.

38 "Operating Experience and Design Features of Closed-Cycle Gas-Turbine Power Plants," by Curt Keller, Trans. ASME, vol. 79, 1957, pp. 627–643.

39 (a) "Research on Application of Cooling to Gas Turbines," by J. B. Esgar, J. N. B. Livingood, and R. O. Hickel, Trans. ASME, vol. 79, 1957, pp. 645–651—paper contains list of subject publications;

(b) "Application of Internal Liquid Cooling to Gas-Turbine Rotors," by Sumner Alpert, R. E. Grey, and D. D. Drake, Trans. ASME, vol. 78, 1956, pp. 1257–1266.

40 "By-Pass Assessment," *Flight*, vol. 71, February 8, 1957, pp. 183–186.

41 "Experimental Cooling of Radial Flow Turbines," by E. N. Petrick and R. D. Smith, ASME Paper No. 54—A-245.

42 "Concepts on Turbojet Engines for Transport Application," by Abe Silverstein and N. D. Sanders, SAE Paper No. 727, 1956.

43 "Über Mischungsvorgänge in gasgefeuerten Brennkammern," by Fritz Schoppe, *VDI-Forschungsheft* 456, Ausgabe B, Band 1956.

44 "Atomisation of Water by Spinning Discs," by J. R. Pattison and J. D. Aldridge, *The Engineer*, vol. 203, April 5, 1957, pp. 514–519 —correlation between droplet dimensioning and liquid and disk parameters.

45 "Die Verwendung schwerer Brennstoffe in Gasturbinen," by H. A. Havemann and K. Mahadevan, *Motortechnische Zeitschrift*, vol. 18, March, 1957, pp. 81–85; April, 1957, pp. 111–112.

46 For schematic, see: "The 200 HP Boeing Model 502 Gas Turbine," by S. D. Hage, *Interavia*, June, 1949; and also "Nernst

Turbine"; "Steam and Gas Turbines," by A. Stodola, McGraw-Hill Book Company, Inc., New York, N. Y., 1927.

47 See, e. g., *Aviation Research and Development*, April, 1957.

48 (a) "Chemical Fuels," *Aviation Week*, vol. 65, November 5, 1956, p. 28;

(b) "High Energy Fuels," by J. R. Cracknell, *Flight*, vol. 71, March 15, 1957, pp. 332–334.

49 See, e.g., Brown Boveri & Company, Ltd.; D. Napier & Son, Ltd.; and Gerätebau Eberspächer OHG Brochures.

50 "Vibration in Gas Turbine Engines," by H. C. Hill, *SAE Journal*, vol. 64, October, 1956, p. 40.

51 "Small Gas Turbine Development," by H. C. Hill, paper presented at Symposium on "The Role of the Small Gas Turbine," October 1, 1955, sponsored by Polytechnic Institute of Brooklyn, Brooklyn, N. Y.

52 See reference (27) a and b; and "Engineering Details of Ford Rotary Heat Exchanger for Gas Turbine Engines," by W. Wai Chao, *Automotive Industries*, vol. 112, May 15, 1955, pp. 54–56.

"Heat Exchangers for Small Regenerative Gas Turbines," by W. Wai Chao, *SAE Journal*, vol. 64, October, 1956, pp. 38–39.

53 "Rover 28/100 Gas Turbine," *The Autocar*, March 15, 1957.

54 "The Development of the Ljungström Steam Turbine and Air Preheater," by F. Ljungström, Proceedings of The Institution of Mechanical Engineers, vol. 160, 1949, pp. 211–223.

55 "Seal Leakage in the Rotary Regenerator and Its Effect on Rotary-Regenerator Design for Gas Turbines," by D. B. Harper, Trans. ASME, vol. 79, 1957, pp. 233–245.

56 "The Turbo-Regenerator as Applied to Gas Turbines," by W. Hryniszak, Proceedings of the General Discussion on Heat Transfer, The Institution of Mechanical Engineers, 1951, pp. 460–464.

57 Dr. H. L. Ritz, C. A. Parsons & Company, Limited, England. See, e.g., "Summary of the Theory and Construction of a Rational Gas Turbine," by L. Ritz, RME56J18a, Distribution Center, Langley Field, Va.

58 Nuclear energy-fired gas turbines and related heat exchanger problems, although very similar, could not be included in this paper.

59 Detailed information in INCO brochures.

60 "Comparative Correlation of the Elevated Temperature Properties of Various Super Alloys," Report No. 68, January 15, 1957, Universal Cyclops Steel Corporation, Bridgeville, Pa.

"High Temperature Properties of Vacuum Melted Super Alloys," by F. M. Richmond, SAE Paper No. 90, 1957.

61 Brochure; Hastelloy Alloy R-235. For illustration of precision cast turbine wheels see, e.g., *Scientific American*, vol. 196, May, 1957, p. 37.

62 "Low Cost Gas-Turbine Blades, a Must for Automotive Use," by D. N. Frey, *Automotive Industries*, vol. 114, January 1, 1956, pp. 52–53, 109–110.

63 "Automatic Ultrasonic Inspection," *Automotive Industries*, vol. 111, July 15, 1954, pp. 49 and 104.

64 "Successful Containment of High-Speed Turbine Failures," by W. W. Houghton and E. R. Phillips, SAE Paper No. 95, 1957.

65 "High-Temperature Brazing Proves OK for Jet Engines," by J. V. Long, G. D. Cremer, and R. S. Mueller, *SAE Journal*, vol. 65, March, 1957, pp. 43–49.

66 "Refractory Nozzle Blades for Gas Turbines," *Gas and Oil Power*, vol. 52, March, 1957, pp. 63 and 65.

67 "100-HR Gas-Turbine Operation," *Mechanical Engineering*, vol. 76, 1954, p. 1008.

68 "New Materials for Small Gas Turbines," by J. H. Harwood, *SAE Journal*, vol. 64, October, 1956, pp. 41–43.

69 Taken from a letter of Mr. S. D. Hage, Chief Engineer, Industrial Products Division, Boeing Airplane Company, to the author.

Discussion

S. D. HAGE.[3] The author has made a very valuable contribution to the literature. The effort expended in reviewing such a vast field of material must be apparent to all. It should also be recognized that this selection of topics and their analysis represents a publication of the wisdom and judgment of one whose experience qualifies him to point out significant thought stimulators to all of us.

To some, the words "small gas turbines" may simply imply a field of theory that has been over-discussed. The potential and

[3] Chief Engineer, Industrial Products Division, Boeing Airplane Company, Seattle, Wash. Assoc. Mem. ASME.

probable value to mankind must not be treated lightly. Mechanical design development and manufacturing process development are never a revolution. Perhaps one has been expected and the evolution that must occur is a disappointment to many. However, unless theories are well conceived, and much discussed, evolution will not be guided towards a useful objective.

There are many thought stimulators in this paper. This discusser will take this opportunity to develop the author's statements regarding the cube-square law and specific speed. These two parameters are much used to develop useful objectives for the guidance of evolution.

The cube-square law tells us that if exact geometric similarity is maintained for a complete engine, a reduction of all linear dimensions to $1/10$ will reduce shaft horsepower to $1/100$ and weight to $1/1000$ while shaft speed will increase to 10/1. The useful result is a reduction in weight per horsepower to $1/10$. Reducing some part dimensions and some clearances tenfold certainly introduces manufacturing difficulty. The degree to which this measure can be applied depends almost entirely on design capability. With maximum exploitation of die forming of parts the result appears to be worth the effort. A completely fresh look at providing the functions normally supplied by the accessory group is mandatory. The picture painted here implies a very large investment in accurate die work and automatic finishing, inspection, and assembly machinery. The economics of such a proposition demand high volume. It seems, however, that the appetite of our society for portable power is almost limitless, both as to the number of specialized items and also the amount of power that can be packed in a small item.

A second obstacle to scale reduction involves the characteristics of the working fluid. First consider that surface roughness and clearance leakage are strictly involved in the manufacturing techniques mentioned before. Potential gain will be lost if roughness and clearance are not held in linear scale. In addition, the fluid characteristics measured by Reynolds number are against us. The gradient here, however, is not steep. A reduction in scale to $1/10$ which could multiply weight per horsepower ratio by $1/10$ will cause an increase in the losses that are measured by Reynolds number of about 50 per cent unless true laminar flow results. The occurrence of true laminar flow, even in a very small turbo machines seems unlikely. A 50 per cent increase in wetted surface skin friction is considerable but hardly enough to overpower the tenfold potential gain in weight per horsepower. It must be remembered, however, that this phenomenon of increase in boundary friction with reduced scale is a fluid characteristic which we cannot change. Whatever the magnitude of its influence, we must attempt to design around it, not change it, unless we can find a "trade name" detergent which will increase the unit mass of air or reduce its viscosity. Specific speed is a much used parameter against which we judge component efficiency. Where Reynolds number is a valuable measure of the effect of scale on boundary friction losses, specific speed is a valuable measure of the component shape and the resulting component efficiency that our combined technology has produced in the past for any combination of flow and head desired. Low specific speed pumps and compressors usually are radial in flow with a large ratio of discharge over inlet diameter in the rotor. High specific speed pumps and fans are usually axial in flow with a large ratio of blade tip over blade hub diameter in the rotor. Rotors used for intermediate specific speeds grade from one shape to the other in proportion to specific speed.

The study of channel flow has taught us that pressure loss is proportioned to the stream dynamic pressure relative to the bounding surface, the channel length, the channel perimeter, and the inverse of the channel cross-section area. Pressure rise in a radial compressor is proportional to the dynamic pressure

equivalent of the rotor tip speed. It seems reasonable then, once tip speed is fixed by the desired pressure rise, to find a rotor and casing configuration for the flow desired that results in the following: The lowest relative velocities between fluid and bounding surfaces; the shortest flow path through the machine; the smallest possible channel perimeters; and the largest possible channel section areas. Enough blade area must be provided to produce the tangential forces which produce the desired pressure rise. All other wetted surface must be minimized in accordance with the foregoing.

It seems to the discusser that low specific speed radial compressors have poor efficiency because of long flow path, large wetted perimeter, and small channel section. High specific speed axial fans would appear to be just the opposite and should therefore have high efficiency. It seems that the reason they don't could result from excessively large friction relative to work done at the blade tip because the pressure rise is limited by the very much lower blade root velocity. There appears to be a region of very limited data between the most efficient radial machines and the most efficient axial machines. The best efficiencies seem to have been obtained in liquid pumps in this intermediate region. This result also seems to stand up under the foregoing analysis. Consider a radial machine with large inlet diameter relative to rotor tip diameter and large absolute tangential velocity relative to rotor tip speed. Such a machine for given rotor tip speed and any flow could provide large fat channels of short length and high work input relative to the boundary surface relative velocities.

It seems to the discusser that other considerations than maximum efficiency have determined the configurations that are most common along the specific speed scale. In most cases probably either the cheapest or the smallest design for the given conditions of quantity and head along with an economical marriage to the power and speed of the driving shaft have been in control. It also seems that the best efficiency can be obtained by a configuration that is between the best efficiency radial machine and the best efficiency axial machine. Lastly, when this optimum efficiency configuration is apparent, it can be used for any combination of head and flow provided shaft speed and rotor diameter can be chosen to suit.

Shaft speed does not appear to have a limit; therefore, the best efficiency configuration can be used for small machines. This permits small machines to benefit most from the cube-square law. If the influence of the cube-square law is to be overcome when going up in size, large machines will be diameter limited. This requires the use of configurations of high flow per unit of face area and seems to trend away from the optimum efficiency configuration previously proposed.

Perhaps these thoughts are idealistic or in error. If so, this discusser will welcome criticism. In any event, the author will have stimulated some thinking towards the proper guidance of evolution.

Author's Closure

The author is grateful for this discussion by one of the pioneering engineers in this field. All of us who have brought a type of engine into being and production which, only ten years ago, was considered to hold little, if any, promise probably agree in essence with the discusser's thoughts. As regards "miniaturization" and its advantages and problems as well as a better assessment of specific speed characteristics, Reynolds number-efficiency relations and certain manufacturing aspects, very rapid practical developments by a few, and lack of basic and applied research seem to explain why, e.g., we cannot yet delineate clearly best configuration and/or size brackets. We do know, though, that good engineering progress has been and is being made and that even tiny gas turbines are in demand for new applications.

Formulations for the Thermodynamic Properties of Steam and Water

By H. C. SCHNACKEL,[1] SCHENECTADY, N. Y.

The thermodynamic properties of steam and water as presented in the well-known Keenan and Keyes' Steam Tables are expressed as formulas and condensed tables. This is done to facilitate the calculation of steam and water properties on high speed digital computers. Methods of iteration and interpolation using these formulas and tables are discussed. Errors incurred in using the formulations as well as the ranges involved are also described.

Nomenclature

c_p = specific heat at constant pressure, Btu/lb/deg F
e = 2.71828.....
h = enthalpy, Btu/lb
H = enthalpy, int atmos $\times$ cm³/gm
p = pressure, lb/in², absolute
P = pressure = $p/14.6959$ = int atmos, absolute
R = perfect gas constant, Btu/lb/deg F
s = entropy, Btu/lb/deg F
S = entropy, int atmos $\times$ cm³/gm/deg C
t = temperature, deg Fahrenheit

T = temperature, deg Kelvin = $255.38 + \dfrac{5}{9} t$

τ = $\dfrac{1}{T}$ = reciprocal thermodynamic temperature

v = specific volume, ft³/lb
V = specific volume, cm³/gm

Superscripts

$'$ = approximate value

Subscripts

f = saturated liquid property
g = saturated vapor property
l = compressed liquid property

Introduction

With the advent of high speed stored program digital computers there has been an increase in the demand for automatic execution of routine calculations. In the steam power industry probably the most frequently performed single task is that of obtaining some property of steam or water. Formulations and methods suitable for automatic computer calculation of the thermodynamic properties of steam and water are presented in this paper.

The three principal regions, i.e., superheated steam, saturated steam and water, and compressed liquid, will be discussed separately. The well-known publication by Keenan and Keyes (1)[2] is

assumed to be the final authority for the accuracy of the formulations.

Superheated Steam Region

The properties v, h, and s are well formulated by Keenan and Keyes (1). These formulas expressed in a form convenient for automatic computation are as follows

$$v = 0.0160185 \left(\frac{4.55504T}{P} + B \right) \dots\dots\dots\dots [1]$$

$$h = F + 0.043557 \left[F_0 P + \frac{B_0}{2} \left(\frac{P}{T} \right)^2 \right.$$
$$\left. \left\{ -B_6 + B_0 \left[B_2 - B_3 + 2B_7 \frac{B_0}{2} \left(\frac{P}{T} \right)^2 \right] \right\} \right] \dots\dots [2]$$

$$s = 0.809691 \log_{10} T - 0.253801 \log_{10} P + 0.00018052T$$
$$- \frac{11.4276}{T} - 0.355579 - 0.0241983\beta \dots\dots [3]$$

where

$$B = B_0 \left[1 + \frac{B_0 P}{T^3} \left\{ B_2 - B_3 + \frac{B_0 P}{T^3} (B_4 - B_5) B_0 P \right\} \right]$$

$$B_0 = 1.89 - B_1$$

$$B_1 = \frac{2641.62}{T} 10^{80970/T^2}$$

$$B_2 = 82.546$$

$$B_3 = \frac{162460}{T}$$

$$B_4 = 0.21828T$$

$$B_5 = \frac{126970}{T}$$

$$B_6 = B_0 B_1 - 2F_0(B_2 - B_3)$$

$$B_7 = 2F_0(B_4 - B_5) - B_0 B_5$$

$$F_0 = 1.89 - B_1 \left(\frac{372420}{T^2} + 2 \right)$$

$$F = 775.596 + 0.63296T + 0.000162467T^2 + 47.3635 \log_{10} T$$

$$\beta = \frac{1}{T} \left[(B_0 - F_0)P + \frac{B_0}{2} \left(\frac{P}{T} \right)^2 \right.$$
$$\left. \left\{ B_6 + \frac{1}{2} B_0 \left(\frac{P}{T} \right)^2 B_0 \left[B_0(B_4 - B_5) - 2B_7 \right] \right\} \right]$$

It must be noted that the high order term of the corresponding formulas of reference (1) is missing from Equations [1], [2], and [3]. For most applications the region influenced by this term, as seen in Fig. 1, is rarely encountered. As a generalization it may be said that appreciable errors are experienced only at entropy levels less than about 1.4. The form of the high order term is shown in Appendix 1 and may be added to the equations if necessary.

Many applications require properties as functions of two variables other than pressure and temperature such as

$$s, v, \text{ or } t = f(p, h)$$
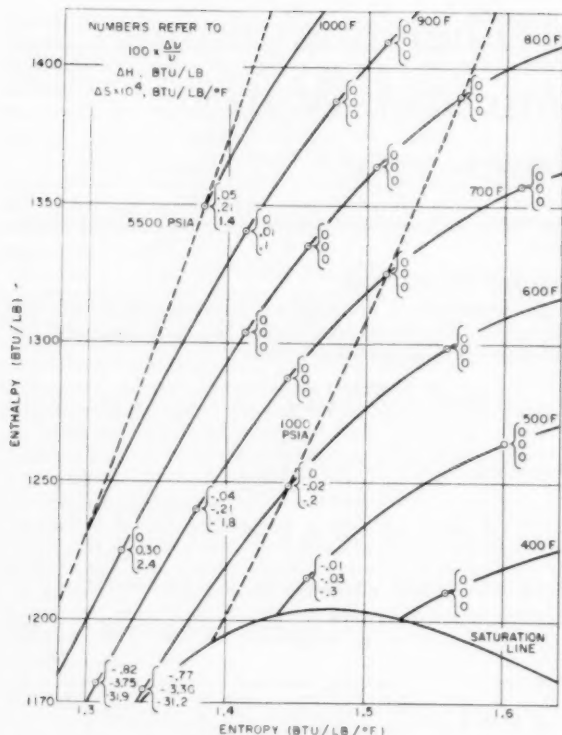
Fig. 1 Variation of high order term in Keenan and Keyes' equations for v, h, and s

$$h, v, \text{ or } t = f(p, s)$$

$$p, t = f(h, s)$$

The case of a temperature being known in the dry region with the pressure unknown occurs so infrequently in practice that it is omitted here. In the case where pressure is known iteration is performed to obtain a temperature. This temperature, then, together with the given pressure, is used in Equations [1], [2], or [3] to obtain the desired property.

Iteration for temperature as a function of pressure and enthalpy requires an initial approximation for the temperature. This is calculated from the following empirical equation

$$t' = a_0 + a_1 h + a_2 h^2 + \frac{p}{b_0 + b_1 h + b_2 h^2} \quad \ldots \ldots [4]$$

where

$a_0 = -2827.7$
$a_1 = 3.0391$
$a_2 = -0.00035413$
$b_0 = 24.318$
$b_1 = -0.071479$
$b_2 = 0.000045305$

The approximate temperature obtained from Equation [4] may be different from the true value by 20 to 30 degrees. However, this difference is quickly reduced by the iteration process. The reciprocal of the specific heat, another quantity required for the iteration process, must be calculated only once. The following formula was found to be satisfactory

$$\frac{1}{c_p} \approx a_1 + 2a_2 h - \frac{p(b_1 + 2b_2 h)}{(b_0 + b_1 h + b_2 h^2)^2} \quad \ldots \ldots [5]$$

After an approximate enthalpy is computed from

$$h' = f(p, t') \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots [6]$$

using Equation [2], a temperature correction is calculated from

$$\Delta t = \frac{1}{c_p}(h - h') \quad \ldots \ldots \ldots \ldots \ldots [7]$$

and the improved temperature becomes

$$t_{\text{improved}} = t' + \Delta t \quad \ldots \ldots \ldots \ldots \ldots [8]$$

The iteration process for $t = f(p, h)$, then, consists of repeatedly solving Equations [6], [7], and [8] until Δt is within a prescribed tolerance.

Iteration for temperature as a function of pressure and entropy also requires the initial calculation of an approximate temperature and reciprocal specific heat. The approximate temperature is given by

$$t' = \alpha_0 + \alpha_1 z + \alpha_2 z^2 \quad \ldots \ldots \ldots \ldots [9]$$

where

$\alpha_0 = 203$
$\alpha_1 = 477.3 + 70.00 \, s$
$\alpha_2 = 440.0 - 24.86 \, s$
$z = 0.57477 \log_{10} p - 3.2949 + 0.93551 \, s + 0.32531 \, s^2$

and the reciprocal of the specific heat may be computed from

$$\frac{1}{c_p} \approx -9.513 + 11.342 \, s - 2.736 \, s^2 \quad \ldots \ldots [10]$$

With the approximate temperature and the given pressure an approximate entropy

$$s' = f(p, t') \quad \ldots \ldots \ldots \ldots \ldots [11]$$

is calculated using Equation [3]. A temperature correction (see Appendix 2) may now be calculated from

$$\Delta t = -(t' + 460) \frac{\Delta s}{c_p}\left[1 + \frac{\Delta s}{2c_p}\right] \quad \ldots \ldots [12]$$

where

$$\Delta s = s' - s$$

Applying this temperature correction an improved temperature is obtained using

$$t_{\text{improved}} = t' + \Delta t \quad \ldots \ldots \ldots \ldots \ldots [8]$$

Thus it is seen that the iteration process for $t = f(p, s)$ involves repeated solutions of Equations [11], [12], and [8] until Δt is within a prescribed tolerance.

For the case where enthalpy and entropy are the independent variables the iteration procedure must involve both temperature and pressure. An approximate pressure may be calculated from the following expression

$$p' = e^{A_1 + A_2} \quad \ldots \ldots \ldots \ldots \ldots [13]$$

where

$A_1 = 9.25(1.9 - s)$
$A_2 = -26.06526 + 0.034723h - 9.48665 \times 10^{-6}h^2$

Determination of an approximate pressure permits the following sequence of calculations to be performed

Table 1

$\log_2 p_0 \qquad x_0 = 0 \qquad y_0$

$$\Delta_0{}^1 y = y_1 - y_0$$

$\log_2 p_1 \qquad x_1 = 1 \qquad y_1$

$$\Delta_0{}^2 y = \frac{\Delta_1{}^1 y - \Delta_0{}^1 y}{2}$$

$$\Delta_1{}^1 y = y_2 - y_1$$

$$\Delta_0{}^3 y = \frac{\Delta_1{}^2 y - \Delta_0{}^2 y}{3}$$

$\log_2 p_2 \qquad x_2 = 2 \qquad y_2$

$$\Delta_1{}^2 y = \frac{\Delta_2{}^1 y - \Delta_1{}^1 y}{2}$$

$$\Delta_2{}^1 y = y_3 - y_2$$

$\log_2 p_3 \qquad x_3 = 3 \qquad y_3$

$\qquad\quad . \qquad\qquad . \qquad\quad .$

$\qquad\quad . \qquad\qquad . \qquad\quad .$

$\log_2 p_k \qquad x_k = 5 \qquad y_k$

Table 2

$$0.125 \leqq p \leqq 2048$$

p	$\log_2 p$	s_f	s_g	h_f	h_g	pv_g	$t_{sat.}$	v_f
0.125	-3.000	0.01810	2.15616	9.04	1079.67	297.93	40.96	0.016020
0.177	-2.500	0.03566	2.12689	17.90	1083.55	303.24	49.82	0.016029
0.250	-2.000	0.05420	2.09702	27.36	1087.70	308.82	59.30	0.016041
0.354	-1.500	0.07306	2.06696	37.27	1092.03	314.64	69.23	0.016053
0.500	-1.000	0.09239	2.03720	47.60	1096.41	320.70	79.58	0.016080
0.707	-0.500	0.11226	2.00763	58.39	1101.13	327.03	90.40	0.016103
1.000	-0.000	0.13260	1.97821	69.70	1106.00	333.59	101.74	0.016140
1.414	0.500	0.15344	1.94894	81.55	1110.97	340.37	113.62	0.016181
2.000	1.000	0.17488	1.91998	93.98	1116.22	347.46	126.08	0.016228
2.828	1.500	0.19698	1.89117	107.07	1121.66	354.85	139.18	0.016291
4.000	2.000	0.21977	1.86246	120.86	1127.31	362.52	152.97	0.016357
5.657	2.500	0.24315	1.83405	135.41	1133.20	370.48	167.51	0.016433
8.000	3.000	0.26764	1.80573	150.79	1139.30	378.72	182.86	0.016528
11.314	3.500	0.29239	1.77768	167.09	1145.50	387.29	199.10	0.016630
16.000	4.000	0.31847	1.74973	184.42	1151.94	396.00	216.32	0.016744
22.627	4.500	0.34531	1.72195	202.88	1158.67	405.00	234.60	0.016884
32.000	5.000	0.37329	1.69415	222.59	1165.33	414.08	254.05	0.017040
45.255	5.500	0.40240	1.66647	243.72	1172.11	423.19	274.79	0.017214
64.000	6.000	0.43282	1.63867	266.45	1178.80	432.15	296.94	0.017420
90.510	6.500	0.46460	1.61069	290.98	1185.36	440.79	320.67	0.017666
128.000	7.000	0.49799	1.58249	317.57	1191.45	448.89	346.13	0.017981
181.019	7.500	0.53308	1.55368	346.52	1196.95	455.83	373.52	0.018280
256.000	8.000	0.57016	1.52426	378.27	1201.34	461.25	403.05	0.018684
362.039	8.500	0.60963	1.49356	413.27	1204.11	464.26	434.95	0.019136
512.000	9.000	0.65166	1.46112	452.23	1204.28	463.68	469.46	0.019804
724.077	9.500	0.69705	1.42599	496.06	1200.61	457.92	506.88	0.020600
1024.000	10.000	0.74660	1.38681	546.07	1190.91	444.32	547.50	0.021662
1448.154	10.500	0.80208	1.34039	604.96	1170.81	418.36	591.56	0.023288
2048.000	11.000	0.86670	1.28009	677.27	1131.55	371.34	639.19	0.025947
2896.309	11.500	0.94749	1.19150	772.00	1054.00	283.05	690.35	0.032840

$$2048 < p \leqq 3000$$

p	$\log_2 p$	s_f	s_g	h_f	h_g	pv_g	$t_{sat.}$	v_f
2048.000	11.000	0.86690	1.28015	677.27	1131.56	371.34	639.19	0.025842
2233.359	11.125	0.88538	1.26111	698.68	1116.28	354.39	651.64	0.026995
2435.496	11.250	0.90594	1.23941	722.69	1097.63	333.06	664.28	0.028247
2655.927	11.375	0.92929	1.21375	750.29	1073.72	308.70	677.10	0.030041
2896.309	11.500	0.95821	1.17987	784.83	1039.67	275.21	690.07	0.032820
3158.447	11.625	1.00055	1.12628	834.58	982.02	224.28	703.09	0.038595

$$t' = f(p', h) \ldots\ldots\ldots\ldots\ldots\ldots [14]$$

$$s' = f(p', t') \ldots\ldots\ldots\ldots\ldots\ldots [15]$$

and

$$\Delta s = s' - s \ldots\ldots\ldots\ldots\ldots\ldots [16]$$

Finally an improved pressure is computed from

$$p_{\text{improved}} = p' e^{9.25 \Delta s} \ldots\ldots\ldots\ldots [17]$$

Equations [14] through [17] are repeated in an iterative manner until Δs is reduced to a predetermined tolerance.

It is of interest to note that in Equation [13] A_2 is a polynomial representation of $\ln p$ at $s = 1.9000$ and applies over the enthalpy range $1118 \le h \le 1618$. In addition A_1 is an approximation to the natural log of the pressure ratio, at constant enthalpy, between the given entropy and an entropy of 1.9000. A short derivation of [17] is shown in Appendix 2.

Saturated Liquid and Vapor Region

The saturation properties for which a method of calculation is presented are t_{sat}, v_f, $p v_g$, h_f, h_g, s_f, and s_g. The method is a fifth order Newtonian table interpolation with $\log_2 p$ as the argument. Interpolation accuracy is unaffected by the base of the logarithmic argument and, since an IBM 704 computer was used, a base of two seemed most convenient.

A brief review of Newton's interpolation formula (3) is in order here. The following discussion is directed toward simplicity in computer programming.

The fifth order Newtonian interpolation polynomial may be written as

$$y = y_0 + x\{\Delta_0{}^1 y + (x - 1)[\Delta_0{}^2 y + (x - 2) \\ \{\Delta_0{}^3 y + (x - 3)[\Delta_0{}^4 y + (x - 4)\Delta_0{}^5 y]\}]\} \ldots \ldots [18]$$

where, in our present situation

$$y = f(p) = \text{desired property}$$

$$x = \frac{\log_2 p - \log_2 p_0}{\Delta \log_2 p}$$

$$\Delta \log_2 p = \log_2 p_1 - \log_2 p_0 = \log_2 p_2 - \log_2 p_1 = \log_2 p_n - \log_2 p_{n-1}$$

$$\Delta_0{}^n y = \frac{\Delta_1{}^{n-1} y - \Delta_0{}^{n-1} y}{n}, \quad n = 1, 2, 3, 4, \text{ or } 5$$

$$\Delta_0{}^1 y = \frac{y_1 - y_0}{1}, \quad \text{for } n = 1$$

Table 1 shows a few of the leading differences ($\Delta_0{}^n y$) developed. Once Table 1 is completed, the solution of Equation [18] for y is straightforward.

Table 2 of Keenan and Keyes (1), including only the properties discussed herein, was condensed almost ten to one to Table 2. Due to the rapidly increasing changes in saturation properties as the critical pressure is approached, the upper limit was set at 3000 psia and the table divided into two ranges.

The rounding error inherent in table 2 (1) caused some difficulty in obtaining by interpolation suitable values for Table 2. This problem was overcome by using ninth-order orthogonal polynomials (2). Individual small pressure ranges, consisting of at least 10 points each, were fitted with orthogonal polynomials. The values of the property at the desired pressures were then calculated from these polynomials. The result was a properly interpolated value for Table 2 with nearly one more significant figure than in table 2 (1). However, in the final analysis a few of the Table 2 values required revision in order that table 2 (1) be suitably reproduced.

The best accuracy for the range

$$\frac{1}{8} \le p \le 2048 \text{ psia}$$

is obtained where

$$x_2 \le x \le x_4 \ldots\ldots\ldots\ldots\ldots [19]$$

in the interpolation interval

$$x_0 \le x \le x_5 \ldots\ldots\ldots\ldots\ldots [20]$$

An exception to this criterion occurs at the low pressure end of the table where the condition $x_0 \le x \le x_2$ may exist. In this particular case no accuracy is lost since the variations in each of the properties are relatively small.

The part of the Table 2 covering the pressure range $2048 < p \le 3000$ psia consists of six points. Here, however, the interval is taken small enough so that a point anywhere in the range $x_0 \le x \le x_4$ will yield a sufficiently accurate interpolated result.

The magnitude of the error encountered when comparing a Table 2 result with an entry in table 2 (1) is seen in the following tabulation

$$|\Delta t_{\text{sat}}| < 0.02 \qquad |\Delta h_f| < 0.10$$

$$\left|\frac{\Delta v_f}{v_f}\right| < 0.0025 \qquad |\Delta h_g| < 0.10$$

$$|\Delta s_f| < 0.0001$$

$$\left|\frac{\Delta p v_g}{p v_g}\right| < 0.0004 \qquad |\Delta s_g| < 0.0001$$

The error plot of h_g as shown in Fig. 2 is typical of all saturation properties discussed.

Table 3

t	h_f	v_f
32	0.00	0.016022
64	32.05	0.016045
96	63.98	0.016119
128	95.91	0.016237
160	127.89	0.016392
192	159.96	0.016580
224	192.17	0.016799
256	224.58	0.017051
288	257.26	0.017334
320	290.28	0.017654
352	323.74	0.018013
384	357.73	0.018417
416	392.40	0.018875
448	427.88	0.019397
480	464.40	0.020002
512	502.19	0.020719
544	541.65	0.021568
576	583.44	0.022631
608	628.67	0.024015
640	678.60	0.026001
672	738.80	0.029244
704	822.70	0.035000
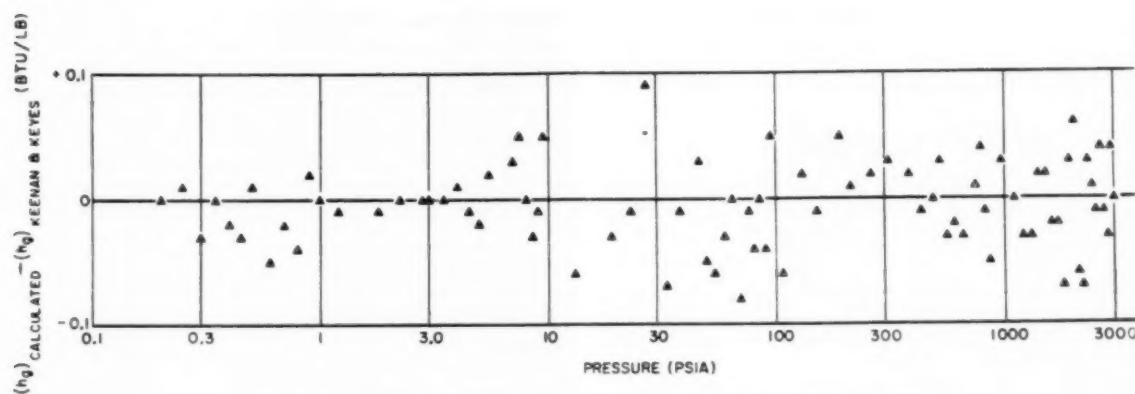
Fig. 2 Error plot of $h_{g\,calc}$ relative to h_g as given by Keenan and Keyes (1)

Table 4 $h_i - h_f$

p \ t	0	100	200	300	400	455	510	565	620
0	0.01	0.00	-0.02	-0.13	-0.27	0.00	0.00	0.00	0.00
500	1.47	1.38	1.09	0.79	0.27	0.00	0.08	2.30	14.00
1000	3.02	2.70	2.21	1.75	0.84	0.15	-0.20	0.70	8.50
1500	4.70	3.99	3.36	2.70	1.44	0.51	-0.50	-1.20	2.47
2000	6.30	5.31	4.51	3.64	2.03	0.79	-0.68	-2.32	-1.80
2500	8.06	6.58	5.63	4.55	2.66	1.09	-0.79	-3.31	-5.60
3000	9.65	7.88	6.76	5.49	3.33	1.53	-0.91	-4.23	-8.70
3500	11.25	9.17	7.90	6.45	4.01	1.99	-0.98	-5.04	-11.20
4000	12.65	10.49	9.03	7.41	4.71	2.43	-0.83	-5.55	-13.30
4500	14.11	11.80	10.15	8.40	5.40	2.87	-0.72	-5.96	-15.10
5000	15.75	13.08	11.30	9.36	6.08	3.25	-0.52	-6.24	-16.70
5500	17.16	14.41	12.47	10.36	6.78	3.70	-0.31	-6.65	-18.00
6000	18.65	15.72	13.62	11.39	7.50	4.23	-0.16	-7.13	-19.10

Table 5 $(v_i - v_f) \times 10^5$

p \ t	0	100	200	300	400	455	510	565	620
0	0.00	-0.19	-0.10	0.00	3.05	12.11	19.55	51.02	190.07
500	-3.00	-2.69	-2.79	-3.50	-3.29	-1.31	4.29	26.80	139.00
1000	-6.10	-5.10	-5.39	-6.89	-8.70	-8.11	-5.79	8.80	77.01
1500	-9.30	-7.51	-8.11	-10.40	-14.09	-16.31	-17.50	-14.00	25.94
2000	-12.35	-9.89	-10.80	-13.80	-19.50	-23.96	-28.99	-35.74	-24.01
2500	-15.14	-12.30	-13.40	-17.19	-24.80	-30.88	-39.60	-53.41	-64.99
3000	-17.91	-14.71	-16.00	-20.69	-29.99	-37.79	-49.71	-69.40	-100.99
3500	-21.20	-16.90	-18.50	-24.20	-35.00	-44.99	-59.10	-83.59	-132.99
4000	-24.01	-19.19	-21.00	-27.49	-40.01	-51.00	-68.21	-98.80	-160.00
4500	-27.11	-21.41	-23.51	-30.71	-44.89	-57.20	-76.72	-111.63	-184.01
5000	-30.30	-23.60	-26.01	-34.00	-49.59	-63.35	-85.31	-124.60	-205.99
5500	-33.31	-25.70	-28.40	-37.19	-54.19	-69.40	-93.65	-136.64	-227.00
6000	-36.19	-27.80	-30.80	-40.51	-58.70	-75.29	-102.21	-151.30	-242.50

Table 6 Errors in liquid properties relative to Keenan and Keyes (1)

$\Delta v = v_{l\,\text{calc}} - v_{l\,\text{Keenan and Keyes}}; \quad \Delta h = h_{l\,\text{calc}} - h_{l\,\text{Keenan and Keyes}}$

p		t	32	100	200	300	400	500	600	620
200.0	Δv x 10^5		-0.2	-0.1	-0.3	-0.2				
	Δh		-0.01	0.01	0.02	0.02				
400.0	Δv x 10^5		-0.1	-0.1	-0.3	0.1	0.1			
	Δh		-0.01	0.01	-0.01	0.00	0.00			
600.0	Δv x 10^5		0.0	0.0	-0.1	0.2	0.0			
	Δh		0.00	-0.03	0.01	0.02	0.00			
800.0	Δv x 10^5		0.1	-0.1	-0.2	0.2	-0.1	-0.4		
	Δh		0.00	0.00	-0.01	0.02	0.00	0.00		
1000.0	Δv x 10^5		0.0	0.0	-0.2	0.1	0.0	0.4		
	Δh		0.00	0.00	0.01	0.01	0.00	0.02		
1500.0	Δv x 10^5		0.0	0.0	-0.2	0.1	0.0	0.6		
	Δh		0.00	0.00	0.01	0.01	-0.01	0.04		
2000.0	Δv x 10^5		0.0	0.0	-0.2	0.1	0.0	0.5	-0.1	0.5
	Δh		0.00	0.00	0.01	0.01	0.00	0.04	-0.01	-0.02
2500.0	Δv x 10^5		0.1	0.0	-0.2	0.1	-0.1	0.5	-0.1	0.5
	Δh		0.01	0.00	0.01	0.00	0.00	0.04	-0.03	-0.02
3000.0	Δv x 10^5		0.1	0.0	-0.2	0.1	-0.1	0.4	-0.1	0.5
	Δh		-0.03	-0.01	0.01	0.01	-0.01	0.03	-0.05	-0.03
3206.2	Δv x 10^5		0.2	0.1	-0.1	0.2	0.0	0.6	0.6	0.3
	Δh		-0.04	-0.04	-0.01	-0.01	-0.01	0.03	-0.05	0.04
3500.0	Δv x 10^5		0.1	0.0	-0.2	0.1	-0.1	0.5	0.1	0.5
	Δh		0.00	0.00	0.01	0.01	0.00	-0.01	-0.05	-0.02
4000.0	Δv x 10^5		0.1	0.0	-0.2	0.1	-0.1	0.5	-0.6	0.5
	Δh		-0.04	0.00	0.01	0.00	0.00	0.03	-0.01	-0.02
4500.0	Δv x 10^5		0.1	0.0	-0.2	0.1	0.0	0.5	0.0	0.5
	Δh		-0.05	0.00	0.01	0.01	0.00	0.02	-0.04	-0.02
5000.0	Δv x 10^5		0.0	0.0	-0.2	0.1	-0.1	0.4	-0.1	0.5
	Δh		0.00	0.00	0.01	0.01	0.00	0.03	-0.03	-0.02
5500.0	Δv x 10^5		0.0	0.0	-0.2	0.1	-0.1	0.3	-0.1	0.5
	Δh		-0.03	0.00	0.01	0.01	0.00	0.04	0.01	-0.02
6000.0	Δv x 10^5		0.0	0.0	-0.2	0.1	0.0	0.4	-0.1	
	Δh		-0.01	0.00	0.01	0.01	0.00	0.03	-0.01	

Compressed Liquid Region

Table interpolation is used to compute the compressed liquid properties, $v_l = f(p, t)$ and $h_l = f(p, t)$. The temperature and pressure ranges are $32 \leq t \leq 620$ F and $0 \leq p \leq 6000$ psia, respectively. The general scheme of calculation involves the addition of the saturated liquid property to the difference between the compressed liquid and saturated liquid properties.

Table 3 shows the saturated liquid properties versus temperature. Again the recommended interpolation procedure is the fifth order Newtonian, with the same restrictions as those imposed by Inequalities [19] and [20]. Because of small changes in v_f and h_f with t, the low temperature end of Table 3 is not held to these restrictions. Most of the values of Table 3 were obtained from Table 2 by inverse interpolation.

Tables 4 and 5 contain $h_l - h_f$ and $(v_l - v_f) \times 10^5$, respectively. To determine the desired difference value two fourth-order Newtonian interpolations in temperature and one linear interpolation in pressure are required. Over the temperature range covered $v_l - v_f$ is less than 10 per cent of v_f and $h_l - h_f$ is never larger in magnitude than 20 Btu/lb. Thus an error in these differences has a greatly reduced effect on the accuracy of the final

compressed liquid property. Therefore no interpolation interval criterion such as [19] was found necessary. It should be noted that interpolation with temperature is done in one of two ranges ($0 \leq t \leq 400$ F and $400 < t \leq 620$ F), the 400 F point being common to both intervals.

Compressed liquid properties computed from Tables 3, 4, and 5 are continuous and agree favorably with properties obtained from table 4 (1). Table 6 shows the errors in h_l and v_l relative to values calculated from table 4 (1).

Concluding Remarks

The direct and iterative formulations for the dry steam properties are easily programmed for a high speed digital computer. Many terms and factors are common among the equations for these properties, thus permitting conservation of computer program storage.

The values in Tables 2, 3, 4, and 5 may, through judicious subtraction, be expressed as five digit numbers. Also the word or number length in most, if not all, commercial computers is 10 decimals digits. Thus two table entries may be combined into one computer word. All the values in Tables 2, 3, 4, and 5 may

therefore be stored in 265 words. These together with a general interpolation program form a compact package.

When this paper was written, the author included information and expressed opinions he believed to be correct and reliable. However, because of the constant advance of technical knowledge, the widely differing conditions of possible specific applications, and the possibility of misapplication, neither the author nor his employer makes any warranty with respect to, or assumes any liability arising out of this paper, its contents, or its use.

Bibliography

1 "Thermodynamic Properties of Steam," by J. H. Keenan and F. G. Keyes, John Wiley & Sons, Inc., New York, N. Y., 1936.
2 "Numerical Calculus," by W. E. Milne, Princeton University Press, Princeton, N. J., 1949.
3 "Practical Analysis," by Dr. Fr. A. Willers, Dover Publications, Inc., New York, N. Y., 1948.

APPENDIX 1

The nomenclature and form of the high order terms omitted from Equations [1], [2], and [3] have been taken directly from Keenan and Keyes (1). A change in units and a few clarifying steps of derivation are included herein.

The high order term for the specific volume equation from reference (1) is

$$\Delta V = \phi_{12} P^{12} \dots\dots\dots\dots [21]$$

where

$$\phi_{12} = B_0 g_2(\tau)[B_0\tau]^{12}$$
$$B_0 = 1.89 - 2641.62\tau 10^{80.9\times\tau^2}$$
$$g_2(\tau) = 3.635 \times 10^{-4} - 6.768 \times 10^{64}\tau^{24}$$

Equation [21] expressed in conventional engineering units of the English system becomes

$$\Delta v = 0.0160185\phi_{12}P^{12} \dots\dots\dots\dots [21a]$$

Since only differences are involved we may write from Maxwell's relations

$$\left(\frac{\partial(\Delta H)}{\partial P}\right)_T = \left(\frac{\partial(\Delta V\tau)}{\partial\tau}\right)_P = \frac{d(\phi_{12}\tau)}{d\tau}P^{12} = F_{12}P^{12} \dots [22]$$

where

$$F_{12} = \frac{d(\phi_{12}\tau)}{d\tau} = -(B_0\tau)^{12}[37B_0g_2 - 0.008724\ B_0$$
$$+ 13g_2(B_0 - 1.89)(1 + 372420\tau^2)]$$

Integration of Equation [22] with respect to P gives

$$\Delta H = \frac{F_{12}}{13}P^{13} \dots\dots\dots\dots [23]$$

which is the form of the high order enthalpy term given in reference (1). In more conventional units Equation [23] becomes

$$\Delta h = 0.043557\left(\frac{F_{12}}{13}\right)P^{13} \dots\dots\dots [23a]$$

Again from Maxwell's relations the following expression is written

$$\left(\frac{\partial(\Delta S)}{\partial P}\right)_T = -\left(\frac{\partial(\Delta V)}{\partial T}\right)_P \dots\dots\dots [24]$$

Changing the variable from T to τ gives for the right side of Equation [24]

$$-\left(\frac{\partial(\Delta V)}{\partial T}\right)_P = \left[\left(\frac{\partial(\Delta V\tau)}{\partial\tau}\right)_P - \Delta V\right]\tau \dots\dots [25]$$

Now combining Equations [21], [22], [24], and [25] results in

$$\left(\frac{\partial(\Delta S)}{\partial P}\right)_T = (F_{12} - \phi_{12})\tau P^{12} \dots\dots\dots [26]$$

Integrating Equation [26] with respect to pressure gives

$$\Delta S = (F_{12} - \phi_{12})\tau\frac{P^{13}}{13} \dots\dots\dots\dots [27]$$

which is identical to the term $\frac{M_{12}}{13}P^{13}$ of equation 16, reference(1).

Equation [27] expressed in conventional units becomes

$$\Delta s = 0.024198\ (F_{12} - \phi_{12})\tau\frac{P^{13}}{13} \dots\dots\dots [28]$$

APPENDIX 2

In this appendix only, T will denote absolute temperature in degrees Rankine.

The following is a derivation of Equation [12], the iteration formula for temperature as a function of entropy at constant pressure. We may write for a constant pressure process

$$T\ ds = c_p dT \dots\dots\dots\dots\dots [29]$$

The first three terms of a Taylor expansion for temperature at an entropy of $(s + \Delta s)$ are

$$T(s + \Delta s) = T(s) + \Delta s\frac{d}{ds}T(s) + \frac{[\Delta s]^2}{2}\frac{d^2}{ds^2}T(s) \dots [30]$$

where Δs is defined as $s' - s$. From Equation [29] the first and second derivatives are obtained, thus

$$\frac{d}{ds}T(s) = \frac{T(s)}{c_p} \dots\dots\dots\dots [31]$$

and

$$\frac{d^2}{ds^2}T(s) = \frac{T(s)}{c_p^2} \dots\dots\dots\dots [32]$$

Equation [30] now becomes

$$T(s + \Delta s) = T(s) + T(s)\frac{\Delta s}{c_p} + T(s)\frac{[\Delta s]^2}{2c_p^2} \dots [33]$$

Factoring $T(s)$ in the right side of Equation [33] and transposing terms gives

$$T(s) - T(s + \Delta s) = -T(s + \Delta s)\left\{\frac{\frac{\Delta s}{c_p}\left[1 + \frac{\Delta s}{2c_p}\right]}{1 + \frac{\Delta s}{c_p}\left[1 + \frac{\Delta s}{2c_p}\right]}\right\}$$
$$\dots\dots [34]$$

Since $\left|T(s) - T(s + \Delta s)\right|$ is, as initially calculated, of the order of 30 F it is seen that

$$\left|\frac{T(s) - T(s + \Delta s)}{T(s + \Delta s)}\right| < 0.05 \dots\dots\dots [35]$$

Therefore we may write with very small error

$$\Delta T = T(s) - T(s + \Delta s) = -T(s + \Delta s)\frac{\Delta s}{c_p}\left[1 + \frac{\Delta s}{2c_p}\right]$$
$$\dots\dots [36]$$

or, as in Equation [12]

$$\Delta t = -(t' + 460)\frac{\Delta s}{c_p}\left[1 + \frac{\Delta s}{2c_p}\right]\dots\dots\dots[12]$$

The derivation of Equation [17], the pressure correction formula for a change in entropy at constant enthalpy, is as follows. We may write the thermodynamic relation

$$T\,ds = dh - v\,dp\dots\dots\dots\dots[37]$$

and for constant enthalpy

$$ds = -\frac{v}{T}\,dp\dots\dots\dots\dots[38]$$

It is to be noted that the terms in Equations [37] and [38] are to have constant units. Introducing the perfect gas law Equation [38] becomes

$$ds = -\frac{R}{p}\,dp\dots\dots\dots\dots[39]$$

Integrating Equation [39] gives

$$s' - s = -R\ln\frac{p'}{p}\dots\dots\dots\dots[40]$$

Assuming a value for R of approximately 84 ft lb/lb/deg F, Equation [40] reduces to

$$p = p'e^{0.\,\mathbf{m}(s'-s)}\dots\dots\dots\dots[41]$$

which is identical to Equation [17].

The Formulation of Steam Properties for Digital Computer Application

By WILLIAM G. STELTZ[1] AND GEORGE J. SILVESTRI,[1] LESTER, PA.

Methods and equations are presented for approximating the properties of steam set forth in the Keenan and Keyes' Steam Tables. Equations and functional relationships for derived thermodynamic properties as functions of basic independent variables are presented for compressed and saturated liquid, saturated and unsaturated vapor, and superheated steam. The thermodynamic properties, enthalpy and entropy, are described throughout while specific volume is determined only in the superheated steam region. Error plots of incremental differences from Keenan and Keyes' values are shown as functions of the independent variables.

Introduction

THE advent of new developments in the digital-computer field has provided the lay engineer with a new and powerful tool applicable to solving even his most complicated problems. Advantageous use of the medium requires the transcription and/or development of methods necessary for solution of his unique problem.

Basic to the field of steam-turbine design and construction is a thorough knowledge and application of the thermophysical prop-

Fig. 1 Temperature-entropy diagram showing regions as defined in this paper

erties of water. Varied terminology may be used to describe water in its many forms as shown in Fig. 1. The several regions and singular lines are defined and referred to throughout this paper. The superheated steam region extends from the saturated vapor line to a maximum temperature of 1600 F ranging between pressures of 1 and 5500 psia. However, equation validity is limited to the domain greater than a fluid specific volume of 0.1603 cu ft per lb. The saturation lines are uniquely defined and are designated liquid and vapor as shown. Pressures higher than the saturation value at a given temperature define a region classified as compressed liquid. Parameters of 6500 psia and 660 F form the limiting boundaries for this region.

Authoritative data, Keenan and Keyes' Steam Tables,[2] have been used as the basis of comparison for determining the accuracy of all approximations. Deviations from this base were evaluated to determine the magnitude of error that could be tolerated and still maintain a realistic level of accuracy in steam cycle calculations.

Error plots are defined as the difference between Keenan and Keyes' values and those calculated by the computer. Acceptable arbitrary limits of error terms were assumed to be 0.10 Btu per lb for enthalpy and 0.0001 Btu per lb deg F for entropy. Property extension beyond the present ranges may be found in the Interim Steam Tables,[3] although adaptation of these data to computer methods was not attempted. However, it is thought that a tabular interpolation method would be the most practical approach in this region.

Methods and Subroutine Designations

Methods of reproducing data on digital computers are relatively few and the appropriate choice must be made in accordance with the peculiarities of a particular problem. Accuracy, memory space required, calculation time, and complexity of the expression are all factors to be considered when a system is to be described.

The most logical method is that of merely reproducing an exact equation describing the dependent variable as a function of the desired independent variables. Equations very often are limited in scope and accuracy over the complete range of variables desired; in this situation, several analytic functions would have to be reproduced. Occasionally, an equation relating not the desired variable but a compatible secondary variable may be available. Through mathematical manipulations it may be possible to express the desired dependent variable in terms of the secondary equation. A high degree of complexity may result from this type of solution, further complicating the problem.

An approximation to the data when an exact equation is not readily available may be practical. This method entails developing an analytical expression over the required range capable of producing results within the desired accuracy. The polynomials, so determined from the curve-fitting methods presently available, ranged (in our systems) from a third-order approxima-
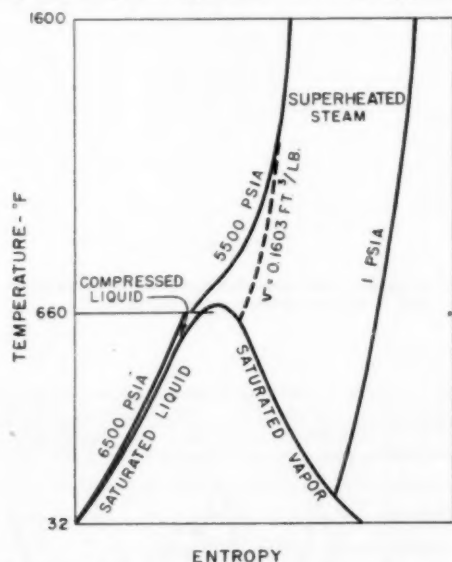
[2] "Thermodynamic Properties of Steam," by J. H. Keenan and F. G. Keyes, John Wiley & Sons, Inc., New York, N. Y., 1955.

[3] "Properties of Steam at High Pressures—An Interim Steam Table," by R. C. Spencer, C. A. Meyer, and R. D. Baird, Mechanical Engineering, vol. 78, 1956, p. 615.

tion to an eighth-order logarithmic fit (derived for TSL[4]—saturation temperature as a function of pressure).

When other methods fail or result in expressions of monstrous complexity, one may resort to a tabular representation of data and an interpolation subroutine. Preparing the table accurately and consistently is often time consuming and tedious work. The interpolation method itself must be chosen to avoid any high-order wiggle or low-order inaccuracy in computed values. A search of routines available through SHARE[5] yielded several possibilities, all of which proved unsuitable for our purposes. A fourth-order LaGrangian interpolation formula was therefore developed for the compressed liquid region, and has proved quite satisfactory. A lower-order routine is sufficient for interpolation in the pressure direction due to nearly linear changes in delta enthalpy values with pressure at constant temperature.

Computer Nomenclature

Symbolic notation for the several calculation methods has been retained for descriptive and identification purposes. Although primarily developed for the IBM 704 digital computer the symbolization was strictly at the programmer's discretion. Three-symbol identification describes all the routines sufficiently and consistently. The initial symbol indicates the dependent variable of the routine, while the remaining two describe the region of system coverage. For instance, the routine PSL describes pressure at saturation; that it is a polynomial function of temperature is not shown and must be obtained from the routine descrip-

Fig. 2 IBM 704 digital-computer subroutine designations

tion. Several of the systems, however, do not rigorously adhere to this identification scheme. For instance, HSS presents enthalpy, entropy, and specific volume in the superheated steam region. The hidden details are not readily apparent, as entropy and specific volume are not indicated by the symbolic notation. The complete list of presented systems is given in Fig. 2. Complete descriptions of each system including ranges, detailed equations, and accuracies are presented here for the reader's convenience.

[4] See Fig. 2.
[5] SHARE, a co-operative organization dedicated to the dissemination of standard IBM 704 subroutines.

Superheated Steam—HSS

As this region is that of the greatest interest in design problems it received a maximum of effort in its derivation and check-out. An exact equation used in the development of the present table is given in the Keenan and Keyes' Steam Tables[2]. Equation (13),[2] specific volume as a function of pressure and temperature, enables us to make use of mathematical relationships to develop equations for thermodynamic properties throughout this entire region, subject of course to various boundary restrictions imposed upon the system.

The present equation is restricted by a minimum specific volume of 0.1603 cu ft per lb where equation validity lies in the specific-volume domain exceeding this minimum value. It would be well to mention that considerable difficulty was encountered in duplicating exactly the results presented in Table 3 of the Keenan and Keyes' Steam Tables. Specific volume and enthalpy correlated excellently, being for the great majority of test points within round-off tolerance. Entropy on the other hand did not possess such ideal characteristics. Errors greater than round-off persisted in the higher-pressure domain with 400 psia seeming to represent an average transition value. The correction equation presented approximately corrects the mean deviation of a given pressure parameter to the zero ordinate value, hence any entropy
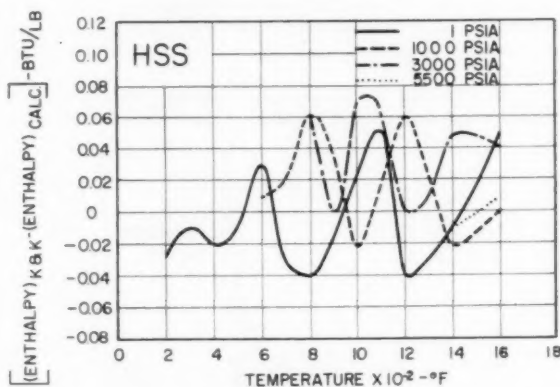
Fig. 3 Plot showing computed enthalpy deviations from Keenan and Keyes' values in the superheated-steam region
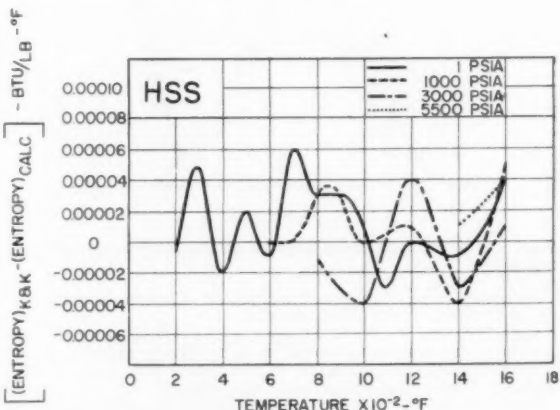
Fig. 4 Plot showing computed entropy deviations from Keenan and Keyes' values in the superheated-steam region

error is that pressure parameter's amplitude about its mean value. Error plots of enthalpy and entropy are presented in Figs. 3 and 4, respectively.

$$\text{HSS} \quad \left.\begin{array}{l}\text{Enthalpy}\\\text{Entropy}\\\text{Specific volume}\end{array}\right] = f(\text{Pressure, Temperature})$$

$$h = [[P\{A - 2\tau\alpha(1 + \beta)\} + \lambda_1\mu_1 + \lambda_2\mu_2 + \lambda_3\mu_3]Q$$
$$+ L\tau^{-1} + \tfrac{1}{2}M\tau^{-2} - N \ln \tau + W + h']U, \quad \text{Btu/lb}$$

$$S = [-\{R \ln P + \tau^2\gamma P + (-\lambda_1 + B_0{}^2\tau^2\delta)\tau\mu_1$$
$$+ (-\lambda_2 + B_0{}^4\epsilon\tau^{-1})\tau\mu_2 - (\lambda_3 + B_0{}^{13}\tau^{-1}\rho)\tau\mu_3\}Q$$
$$- L \ln \tau + M\tau^{-1} - N\tau + S']V + \Delta S, \quad \text{Btu/lb-deg F}$$

where ΔS is an entropy correction term

$$\Delta S = -0.000108 + 0.0000363 \ln (\text{pressure in psia})$$

$$\nu = \left[\frac{R}{P\tau} + B_0 + B_0{}^2\tau^2\delta P + B_0{}^4(\tau P)^3\epsilon - B_0{}^{13}(\tau P)^{12}\rho\right]Y,$$
$$\text{ft}^3/\text{lb}$$

$$\lambda_1 = B_0{}^2\tau^2(3\delta - F\tau) - \delta\tau^3 2B_0\gamma$$
$$\lambda_2 = B_0{}^4\tau^{-1}(4\epsilon - 2H\tau^2) - \epsilon 4B_0{}^3\gamma$$
$$\lambda_3 = -B_0{}^{13}\tau^{-1}[13\rho - 24K(10^2\tau)^{24}] + \rho 13B_0{}^{12}\gamma$$
$$\tau = (K' + T)^{-1}$$

$$\alpha = BC^{D\tau^1} \qquad \rho = J - K(10^2\tau)^{24}$$
$$\beta = D(\ln C)\tau^2 \qquad B_0 = A - \tau\alpha$$
$$\gamma = \alpha[1 + 2\beta] \qquad \mu_1 = P^2/2$$
$$\delta = E - F\tau \qquad \mu_2 = (\tau P)^4/4$$
$$\epsilon = G - H\tau^2 \qquad \mu_3 = (\tau P)^{13}/13$$

$A = 1.89$	$J = 3.635 \times 10^{-4}$	$V = 0.23888888$
$B = 2641.62$	$K = 6.768 \times 10^{-3}$	$W = -698.65$
$C = 10$	$L = 1.4720$	$Y = 0.0160185$
$D = 80870$	$M = 7.5566 \times 10^{-4}$	$C_1 = 14.6959$
$E = 82.546$	$N = 47.8365$	$\ln C = 2.30258509$
$F = 1.6246 \times 10^5$	$Q = 0.101325$	$h' = 2502.36$
$G = 0.21828$	$R = 4.55504$	$K' = 273.16$
$H = 1.2697 \times 10^5$	$U = 0.43$	$S' = -1.48847$

where

$$T = \text{temperature in deg C} = 5/9 (\text{deg F} - 32)$$

$$P = \text{pressure in international atmospheres} = \frac{\text{psia}}{14.696}$$

Saturated Liquid Line
TSL

Temperature as a function of pressure presented a relatively simple curve-fitting problem. A logarithmic polynomial composed of two sections reproduces the dependent variable faithfully up to the critical pressure. A maximum error of approximately 0.05 F occurs at 1500 psia as shown in Fig. 5.

TSL Temperature $= f(\text{Pressure})$

$$0.2 \leqslant P < 450$$
$$T = \Sigma A_i[\ln (10P)]^i$$

$$A_0 = 35.157890$$
$$A_1 = 24.592588$$
$$A_2 = 2.1182069$$
$$A_3 = -0.34144740$$
$$A_4 = 0.15741642$$
$$A_5 = -0.031329585$$
$$A_6 = 0.0038658282$$
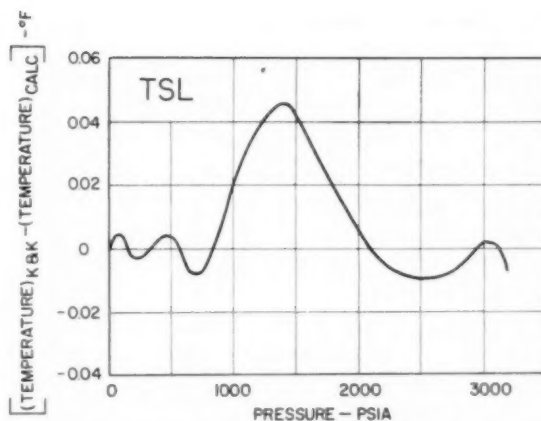$$A_7 = -0.00024901784$$
$$A_8 = 0.0000068401559$$

Fig. 5 Plot showing the computed temperature deviation from Kennan and Keyes' values at saturation conditions

$$450 \leqslant P \leqslant 3206$$
$$T = \Sigma B_i(\ln P)^i$$

$$B_0 = 11545.164$$
$$B_1 = -8386.0182$$
$$B_2 = 2477.7661$$
$$B_3 = -363.44271$$
$$B_4 = 26.690978$$
$$B_5 = -0.78073813$$

where

$$T = \text{deg F}$$
$$P = \text{psia}$$

PSL

Expressions relating the vapor pressure of water and corresponding temperatures have been developed by Smith, Keyes, and Gerry and are presented as equations (11) and (12) in the Keenan and Keyes' Steam Tables. Correlation is again excellent, the maximum error being approximately 0.07 psia at a temperature of 550 F. An error of this magnitude at a lower temperature would be quite serious, but the error becomes insignificant when compared to the saturation pressure at 550 F. The error plot is presented in Fig. 6.
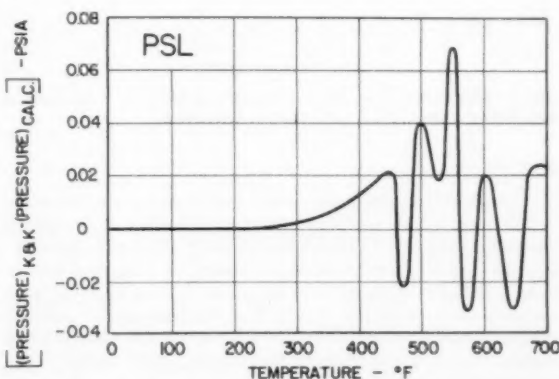

Fig. 6 Plot showing the computed pressure deviation from Keenan and Keyes' values at saturation conditions

PSL   Pressure = f (Temperature)

$$50 \leqslant T < 200$$

$$\text{Log}_{10}\left(\frac{P_c}{P}\right) = \frac{x}{T_k}\left[\frac{A + Bx + Cx^3}{1 + Dx}\right]$$

$A = 3.2437814$
$B = 5.86826 \times 10^{-3}$
$C = 1.1702379 \times 10^{-6}$
$D = 2.1878462 \times 10^{-3}$

$$200 \leqslant T \leqslant 705$$

$$\text{Log}_{10}\left(\frac{P_c}{P}\right) = \frac{x}{T_k}\left[\frac{A + Bx + Cx^3 + Ex^4}{1 + Dx}\right]$$

$A = 3.3463130$
$B = 4.14113 \times 10^{-3}$
$C = 7.515484 \times 10^{-9}$
$D = 1.3794481 \times 10^{-2}$
$E = 6.56444 \times 10^{-11}$

where

$$P = \text{international atmospheres} = \frac{\text{psia}}{14.696}$$

$T_k = \text{deg K} = 5/9 (\text{deg F} - 32) + 273.16$
$P_c = \text{critical pressure} = 218.167 \text{ int atm}$
$T_c = \text{critical temperature} = 647.27 \text{ deg K}$
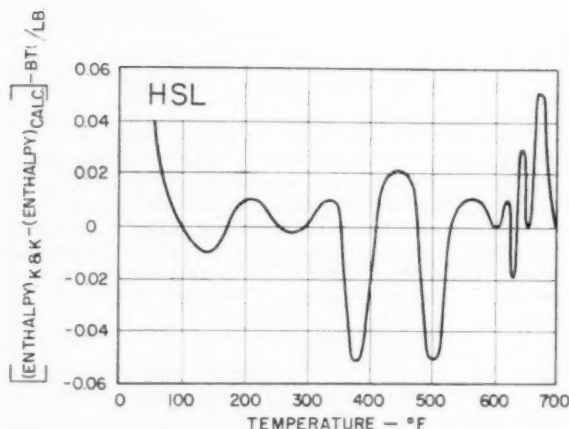$x = T_c - T_k$
$T = \text{deg F}$

$$H = \Sigma A_i(T)^i$$

$$50 \leqslant T < 360$$

$A_0 = -3.2179105 \times 10$
$A_1 = 1.0088084$
$A_2 = -1.1516996 \times 10^{-4}$
$A_3 = 4.8553836 \times 10^{-7}$
$A_4 = -7.3618778 \times 10^{-10}$
$A_5 = 9.6350315 \times 10^{-13}$

$$360 \leqslant T \leqslant 600$$

$A_0 = -9.0411706 \times 10^2$
$A_1 = 1.0673802 \times 10$
$A_2 = -4.2753836 \times 10^{-2}$
$A_3 = 9.41244 \times 10^{-5}$
$A_4 = -1.0315357 \times 10^{-7}$
$A_5 = 4.560246 \times 10^{-11}$

where

$T = \text{deg F}$
$H = \text{Btu/lb}$
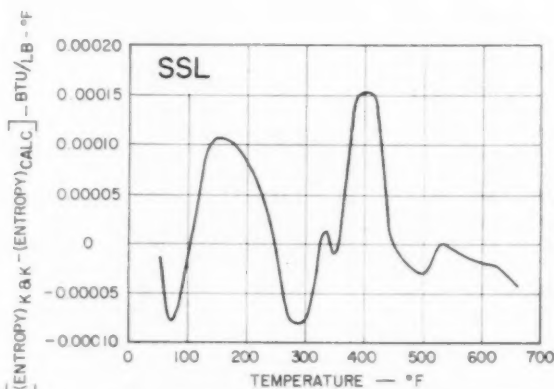
Fig. 8 Plot showing the computed entropy deviation from Keenan and Keyes' values along the saturated-liquid line

SSL

The problems encountered here were similar to those of HSL as was the method of solution. A polynomial composed of two sections suffices to a temperature of 670 F; above this temperature and up to 700 F we must again resort to tabular interpolation methods. The error plot is presented in Fig. 8.

SSL   Entropy = f (Temperature)

$$S = \Sigma A_i(\overline{T})^i$$

$$50 \leqslant T < 450$$

$M = -360 \qquad N = 310$
$A_0 = 5.1575516 \times 10^{-1}$
$A_1 = 3.9679646 \times 10^{-1}$
$A_2 = -4.5979941 \times 10^{-2}$
$A_3 = 3.4251697 \times 10^{-3}$
$A_4 = -6.072333 \times 10^{-3}$
$A_5 = -3.670358 \times 10^{-3}$
$A_6 = 1.2035893 \times 10^{-3}$
$A_7 = 1.234655 \times 10^{-3}$

Fig. 7 Plot showing the computed enthalpy deviation from Keenan and Keyes' values along the saturated-liquid line

HSL

A simple polynomial fit provides the desired accuracy over most of the range for the enthalpy of saturated liquid. High rates of curvature of enthalpy versus temperature in the region approaching the critical point forced abandonment of analytical expressions. Tabular interpolation was instead resorted to in the temperature range from 600 to 700 F. The interpolation method used is essentially identical to that used in the compressed liquid region, that is, a fourth-order LaGrangian interpolation. The error plot is presented in Fig. 7.

HSL Enthalpy = f (Temperature)

$$450 \leqslant T \leqslant 670$$

$$
\begin{aligned}
M &= -560 & N &= 110 \\
A_0 &= 7.6209767 \times 10^{-1} \\
A_1 &= 1.3690825 \times 10^{-1} \\
A_2 &= 7.5137702 \times 10^{-3} \\
A_3 &= 5.7828937 \times 10^{-3} \\
A_4 &= -1.6168801 \times 10^{-3} \\
A_5 &= -2.1403201 \times 10^{-3} \\
A_6 &= 3.9726934 \times 10^{-3} \\
A_7 &= 3.4265601 \times 10^{-3}
\end{aligned}
$$

where

$$
\begin{aligned}
T &= \text{deg F} \\
S &= \text{Btu/lb-deg F} \\
\bar{T} &= \frac{T + M}{N}
\end{aligned}
$$

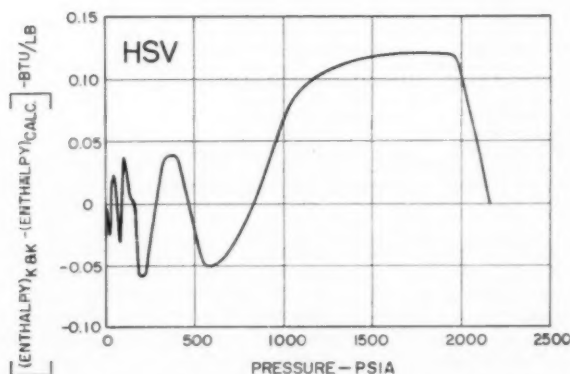Fig. 9 Plot showing the computed enthalpy deviation from Keenan and Keyes' values along the saturated-vapor line
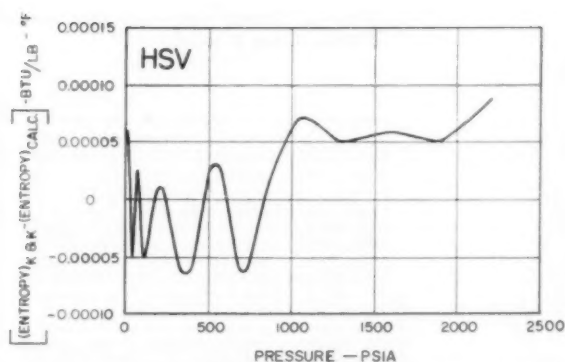

Fig. 10 Plot showing the computed entropy deviation from Keenan and Keyes' values along the saturated-vapor line

Saturated-Vapor Line—HSV

Along the saturated-vapor line both dependent variables were incorporated in one routine. Separate and unique functions comprise each computation although both are defined when the subroutine is specified. A calculation method was devised using the equation derived for HSS. Saturation temperature as a function of pressure (using TSL) may be determined, and then introduced into HSS yielding enthalpy, entropy, and specific volume. This scheme is limited in range to a maximum pressure of 2200 psia due to the decrease in specific volume to values below 0.1603 cu ft per lb. The entropy correction term of HSS, ΔS, is modified for these calculations and is as follows

$$\Delta S = -0.000028 + 0.0000363 \ln (\text{pressure in psia})$$

Rapidly changing functions near the critical point again led to fourth-order LaGrangian interpolation as the easiest and most readily available method in this region. Error plots of enthalpy and entropy are presented in Figs. 9 and 10, respectively.

Table 1 Enthalpy corrections from saturation values in the compressed-liquid region, Btu per lb

HCL

PRESSURE - PSIA	TEMPERATURE - °F																	
	50	100	150	200	250	300	350	400	450	500	550		560	580	600	620	640	660
000	0.02	0.00	-0.02	-0.05	-0.09	-0.15	-0.18	-0.20	-0.10	0.55	2.10							
500	1.48	1.37	1.24	1.10	0.95	0.80	0.56	0.28	0.10	0.13	1.07							
1000	2.93	2.70	2.45	2.21	2.00	1.75	1.38	0.84	0.32	-0.14	0.10							
1500	4.34	3.99	3.64	3.36	3.07	2.70	2.15	1.44	0.65	-0.29	-0.80		-0.70	-0.30	0.60	2.60	8.00	25.10
2000	5.79	5.31	4.90	4.51	4.13	3.64	2.91	2.03	1.00	-0.38	-1.65		-1.85	-2.20	-2.50	-1.80	0.75	11.00
2500	7.20	6.58	6.08	5.63	5.15	4.55	3.70	2.66	1.33	-0.41	-2.40		-2.76	-3.70	-4.90	-5.60	-5.40	-3.10
3000	8.67	7.88	7.29	6.76	6.27	5.49	4.55	3.33	1.71	-0.41	-3.05		-3.60	-5.05	-6.90	-8.70	-10.30	-11.80
3500	10.07	9.17	8.50	7.90	7.26	6.44	5.38	4.01	2.15	-0.34	-3.55		-4.33	-6.30	-8.60	-11.20	-14.30	-17.80
4000	11.45	10.49	9.71	9.03	8.32	7.41	6.24	4.71	2.62	-0.16	-3.95		-4.92	-7.20	-10.00	-13.30	-17.40	-22.50
4500	12.87	11.80	10.95	10.15	9.39	8.40	7.09	5.40	3.14	0.02	-4.30		-5.40	-7.94	-11.10	-15.10	-19.90	-26.40
5000	14.24	13.08	12.13	11.30	10.43	9.36	7.97	6.08	3.61	0.25	-4.60		-5.76	-8.58	-12.10	-16.70	-22.00	-29.50
5500	15.65	14.41	13.49	12.47	11.55	10.36	8.86	6.78	4.15	0.52	-4.85		-6.12	-9.13	-13.10	-18.00	-23.70	-32.00
6000	17.06	15.72	14.66	13.62	12.64	11.39	9.71	7.50	4.68	0.79	-5.05		-6.40	-9.65	-14.00	-19.20	-25.40	-34.20
6500	18.45	17.03	15.85	14.81	13.74	12.40	10.65	8.24	5.16	1.08	-5.23		-6.63	-10.10	-14.60	-20.20	-26.80	-36.30

Table 2 Entropy corrections from saturation values in the compressed-liquid region, Btu per lb deg F $\times 10^3$

HCL

TEMPERATURE - °F

PRESSURE - PSIA	50	100	150	200	250	300	350	400	450	500	550		560	580	600	620	640	660
000	0.03	0.06	0.04	0.03	0.08	0.16	0.31	0.65	1.52	3.30	7.90							
500	0.00	-0.23	-0.42	-0.60	-0.70	-0.75	-0.79	-0.69	-0.25	0.87	4.00							
1000	-0.04	-0.53	-0.90	-1.20	-1.43	-1.64	-1.82	-2.00	-1.94	-1.41	0.31							
1500	-0.08	-0.86	-1.38	-1.79	-2.15	-2.53	-2.92	-3.32	-3.57	-3.56	-2.93		-2.52	-1.40	0.45	3.60	10.20	24.00
2000	-0.17	-1.18	-1.85	-2.39	-2.90	-3.42	-4.00	-4.57	-5.12	-5.58	-5.59		-5.50	-5.08	-4.30	-2.60	1.10	9.00
2500	-0.28	-1.48	-2.33	-2.97	-3.60	-4.25	-4.99	-5.79	-6.65	-7.54	-8.22		-8.33	-8.48	-8.50	-8.20	-6.90	-3.40
3000	-0.39	-1.79	-2.75	-3.56	-4.36	-5.12	-6.02	-7.03	-8.18	-9.42	-10.87		-11.20	-11.80	-12.40	-13.00	-13.40	-13.30
3500	-0.50	-2.08	-3.22	-4.14	-5.02	-5.97	-7.02	-8.21	-9.60	-11.24	-13.31		-13.80	-14.85	-16.00	-17.40	-19.10	-20.90
4000	-0.60	-2.42	-3.73	-4.74	-5.72	-6.77	-8.00	-9.40	-11.05	-13.03	-15.60		-16.20	-17.60	-19.30	-21.40	-24.00	-27.20
4500	-0.72	-2.74	-4.19	-5.33	-6.40	-7.60	-8.99	-10.58	-12.50	-14.80	-17.90		-18.68	-20.35	-22.40	-25.10	-28.30	-32.70
5000	-0.85	-3.07	-4.69	-5.92	-7.09	-8.40	-9.93	-11.74	-13.81	-16.47	-20.09		-20.98	-22.92	-25.30	-28.50	-32.20	-37.50
5500	-0.99	-3.39	-5.10	-6.47	-7.78	-9.20	-10.89	-12.86	-15.21	-18.10	-22.13		-23.06	-25.38	-28.00	-31.50	-35.60	-41.50
6000	-1.20	-3.72	-5.55	-7.06	-8.44	-10.00	-11.83	-13.96	-16.47	-19.57	-24.06		-25.16	-27.70	-30.60	-34.52	-39.05	-45.70
6500	-1.42	-4.05	-6.01	-7.55	-9.11	-10.77	-12.77	-15.06	-17.72	-21.19	-25.90		-27.10	-29.80	-33.10	-37.33	-42.26	-49.50

Compressed Liquid—HCL

Several methods of obtaining the properties of compressed liquid were considered before the final solution evolved as the simplest and most accurate. In the development process an analytic expression was obtained for specific volume and mathematically operated upon to yield desired expressions for the computation of enthalpy and entropy. Accuracy was slightly poorer than desired in the lower-temperature region and larger deviations were experienced as temperature exceeded 500 F.

Due to the complexity and amount of work involved in obtaining an analytical approximation for this region, tabular interpolation methods were investigated. Tabulated values as presented in the Steam Tables are spaced sufficiently far apart to warrant a study of this range pertaining to intermediate values. Enthalpy and entropy therefore were cross plotted and consistent intermediate values obtained. These values are presented in Tables 1 and 2. Standard available interpolation routines were investigated yielding unsatisfactory results. Curve shapes and accuracy requirements led to the development of a fourth-order LaGrangian interpolation routine which is used in both directions of interpolation.

Acknowledgment

The authors wish to express their appreciation to several persons for their efforts in connection with this project. Although individual acknowledgment is impossible, outstanding contributions were made by the South Philadelphia Computer Section, Robert E. Vaillancourt, supervisor, and Dr. Peter A. Zaphyr and R. Lynn Wyatt of the East Pittsburgh Works Analytical Department.

Discussion

Frederick G. Keyes.[6] The application of modern computers to secure material for tabulating thermodynamic or thermophysi-

cal properties generally is a major advance. The Keenan and Keyes' Steam Tables were computed using "Millionaire" computers. However, steam tables covering the range of pressure and temperature to 15000 psia and 1600 F present an almost impossible prospect if no more rapid and error-proof means were available than in the "good old days." The authors have given a comprehensive survey of the extraordinary advantage in the saving of time and assurance of accuracy through the use of the IBM 704 Computer. Incidentally, it is comforting to know the degree of accuracy consistently attained under the system of controls set up at the time the Keenan and Keyes tables were computed from the steam properties experimental data resulting from the older ASME steam research program.

There remains for the desk computer, however adaptable the speedy modern digital computer, the task of correlating the basic data and testing analytical expressions to cover data ranges. This effort results in formulas which, if developed with proper regard to facilitate programming for the new computers, makes possible the production of the new steam tables in a small fraction of the time required formerly. Then too, the accuracy prospects are a welcome feature of the new computers which is difficult to match under the old conditions.

With respect to the limitation of equation 13, page 15 of the Keenan and Keyes' Steam Tables to volumes not less than 0.1603 cf per lb, it should be stated that this defect is inherent in any form so far proposed for pvT properties of the type $v = f(p, T)$. It is possible, however, using the form $p = \phi(v, T)$, to represent the data to volumes less than the critical and this was done some years ago in the case of steam (1949). However, pressure and temperature are the desired variables for engineering purposes and the programming for the digital computers becomes more complicated relative to programming with $v = f(p, T)$ correlating equations.

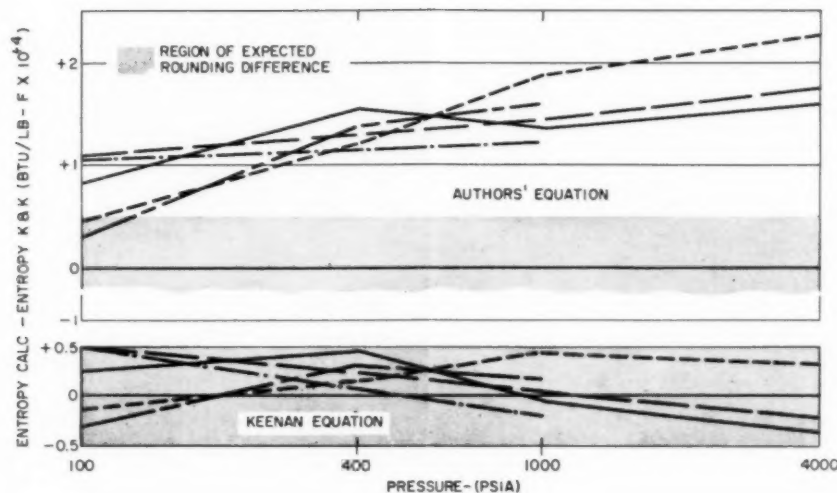**Robert C. Spencer.**[7] The authors are to be commended on pre-

Fig. 11

senting a complex computer subject in a manner easily understood by power engineers. These new computerized versions of an old tool are necessary if the power industry is to continue its leadership in the use of modern business and engineering methods.

It is interesting to note that although Mr. Schnackel[8] indicated good agreement between calculated values using Keenan's equation and Keenan's table, the present authors are unable to obtain similar agreement. We have calculated entropy using the complete Keenan equation and found the deviation from the values in the tables to be well within the expected rounding error as indicated on the bottom half of Fig. 11. The top half of Fig. 11 shows the deviation we have calculated when using the authors' complete equation as given in the paper. In addition, a comparison of the authors' equation for entropy with that of Mr. Schnackel (including the high order term described in the appendix of his paper) shows complete agreement in every respect except for the correction term the authors have added to Keenan's equation.

Therefore, I find it difficult to resolve the apparent discrepancy the authors have found between Keenan's equation and tabular values which has necessitated the addition of a correction term to Keenan's equation.

It would be of interest to know how the authors obtain values of the properties with variables other than pressure and temperature, since such cases are more frequently found in actual practice; for instance, enthalpy as a function of pressure and entropy when calculating an available energy. The methods used for these "indirect" properties are especially important since the computer time involved is much longer and consequently of greater significance than that of the "direct" properties.

The authors mention that a tabular interpolation scheme is used for saturated enthalpy and entropy near the critical point but do not provide the tabular values or an indication of the accuracy obtained. Such information would add considerable value to the paper.

[8] "Formulation of the Thermodynamic Properties of Steam and Water," by H. C. Schnackel, published in this issue, pp. 959–966.

The development of accurate and efficient computer programs for the basic tools of the power engineer are a necessary prerequisite to the use of high speed computers. The authors' paper and those of the other authors presented at the same session are a valuable addition to the arsenal of modern tools for the engineer.

Authors' Closure

The remarks of Messrs. Keyes and Spencer are appreciated. The determination of specific volumes less than 0.1603 ft³/lb using an equation of the form, $p = f(v, T)$ is a possible avenue of attack; however, the method involves iteration to determine the volume. This condition coupled with the fact that the region involved, while of interest to steam generator and heat-transfer apparatus manufacturers, is outside the range of desirable operating conditions for steam turbines. As a consequence, the authors felt that any extensive effort in this region was not warranted.

In the discussion of the superheated steam region routine HSS it was pointed out that good correlation of entropy with Keenan and Keyes' Steam Table values was lacking, hence a correction term was devised to bring the computed values to within an allowable tolerance of the accepted values. With the constants provided in this paper and using the ΔS correction term the system is compatible. The reason for this calculation method is that an incorrect conversion constant has been included. The constant Q should be 0.101295. The use of this number yields results within the working tolerances and thus the correction term ΔS is eliminated. The saturated vapor line routine HSV is also affected as it is a function of HSS.

The basic routines only were presented herein and are functions of the independent variables pressure and temperature. They may be combined with estimates of these variables as functions of one or more of the dependent variables; in conjunction with a standard iterative method the desired functions at a given state point may then be obtained. In the routines where tabular interpolation is indicated, i.e., in the regions approaching the critical point, the values found in the steam tables are used in conjunction with the fourth-order LaGrangian interpolation routine.